Statistical Analysis of Effluent Bioassays

Technical Report
E19

Statistical Analysis of Effluent Bioassays

R&D Technical Report E19

M Crane and E Godolphin

Research Contractor:
Royal Holloway
University of London

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

Publishing Organisation:
Environment Agency
Rio House
Waterside Drive
Aztec West
Almondsbury
Bristol BS32 4UD

Tel: 01454 624400 Fax: 01454 624409

ISBN: 1 85705 382 6

This report is the result of work jointly funded by the Environment Agency and the Scotland and Northern Ireland Forum for Environmental Research (SNIFFER).

Dissemination status
Internal: Released to Regions
External: Released to Public Domain

Statement of use
This report summarises the analysis of effluent bioassay data with different statistical tools. The information within this document is for use by Agency staff and others involved in the use of bioassays for controlling and monitoring discharges to surface waters, particularly within the framework of Direct Toxicity Assessment.

Research contractor
This document was produced under R&D Project E1-003 (Package 8) by:

Royal Holloway
University of London
Egham
Surrey
TW20 0EX

Tel: 01784 443372 Fax: 01784 470756

Environment Agency's Project Manager
The Environment Agency's Project Manager for R&D Project E1-003 was:
Dr Dave Forrow, National Centre for Ecotoxicology & Hazardous Substances, Wallingford.

Package Leader for R&D Project E1-003 Package 8 was: Dr Derek Tinsley, NCEHS.

The SNIFFER Project Manager was: Dr John Redshaw, SEPA West Region, East Kilbride.

R&D Technical Report E19

CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Accelerated failure model A form of **survival time model** (Newman, 1995).

Accelerated life testing A form of **survival time model** (Newman, 1995).

Accuracy The closeness of a measured value to the 'true' value (Newman, 1995).

Acute effect One having a sudden onset or lasting a short time (Rand, 1995).

Acute to Chronic Estimation (ACE) Software package developed by Mayer *et al.* (1996) that includes two-step linear regression and **accelerated life testing** routines for survival data.

Acute-to-chronic ratio (ACR) A numerical unitless value that is the ratio of an acute toxicity test result to a chronic toxicity test result. It is the inverse of the **application factor** (Rand, 1995).

Application factor A numerical unitless value calculated as the threshold chronically toxic concentration of a chemical divided by its acutely toxic concentration (Rand, 1995).

Arcsine transformation A method of transforming data that consist of proportions or percentages so that variances become constant, the distribution becomes normal and parametric statistical methods can be used. The arcsine transformation is the arcsine of the square root of the proportion (Newman, 1995).

Assessment endpoint The entity that a risk assessor wishes to protect (Suter, 1993).

Binomial method A method for deriving an LC50 and confidence interval when there are no partial mortalities (Stephan, 1977).

Bioassay An experiment for estimating the nature, constitution or potency of a material (or of a process), by means of the reaction that follows its application to living matter (Rand, 1995).

Chronic effect Involving a stimulus that is lingering or continues for a long time (Rand, 1995).

Coefficient of determination (r^2) The proportion or percentage of the total variation in the y variable that is accounted for by the fitted line in a linear regression (Zar, 1984).

Concentration-response curve A curve describing the relationship between different exposure concentrations of an agent or material and percentage response of the exposed test population (Rand, 1995).

Correlation coefficient Also known as the simple correlation coefficient, r, the product-moment correlation coefficient, or Pearson's correlation coefficient. It is a numerical, unitless measure of the intensity of association between two variables (Zar, 1984).

Debtox A software package produced by Kooijman and Bedaux (1996d) that allows the time-dependent toxicity of substances to be assessed from acute and chronic data in OECD aquatic toxicity tests with fish, waterfleas and algae. **LC**, **EC** and **NEC** values

	can be obtained from the program.
Degrees of freedom (df)	A quantity found in statistical analysis that takes account of the number of classes of data in the analysis (Zar, 1984). This is in contrast to **stochastic** or probabilistic models.
Deterministic model	A mathematical model that does not account for the uncertainty in the output of the system being modelled (Newman, 1995).
Direct Toxicity Assessment	The total toxic effect of an effluent measured directly with aquatic organisms in a toxicity test. Known as a Whole Effluent Toxicity (WET) test in the United States (Rand, 1995).
Effective concentration (ECx)	The concentration at which an effect of magnitude x occurs. The x is usually 50% of the exposed population, in which case EC50 is known as the median effective concentration (Newman, 1995).
Effluent	A complex waste material that may be discharged into the environment (Rand, 1995).
Elimination rate	The rate at which a bioaccumulated substance is eliminated from an organism by active or passive means (Newman, 1995).
Empirical model	A mathematical model that describes a set of data in a largely theory-free manner (eg a linear regression).
Hazard based model	A mathematical model that uses probability of death per exposure period as a measure of toxic effect (Newman, 1995).
Hormesis	The stimulation at low chemical concentrations of growth, reproduction or some other endpoint in a toxicity test.
Killing rate	A parameter used in **Debtox** (Kooijman and Bedaux, 1996d) and defined as the probability of dying, per unit of time and per unit of concentration that exceeds the **No Effect Concentration**.
Kinetic model	A mathematical model that incorporates knowledge of the movement of chemicals into and out of living organisms. A simple example is a one-compartment first order kinetics model in which the organism is treated as a single 'fully-mixed' compartment and the elimination rate of the compound from the organism is assumed to be exponential (Newman, 1995).
Lethal concentration (LCx)	The concentration at which a lethal effect of magnitude x occurs. The x is usually 50% of the exposed population, in which case LC50 is known as the median lethal concentration (Newman, 1995).
Limit test	A toxicity test in which one concentration is compared with a control. The test concentration will have been selected from several in a previous concentration-response test (Whitehouse *et al.*, 1996).
Logit	The 'log odds unit' used to linearise responses measured in a toxicity test (Newman, 1995).

Lowest Observed Effect Concentration (LOEC) The lowest concentration of a material used in a toxicity test that has a statistically significant adverse effect on the exposed population of test organisms compared with the controls (Rand, 1995).

Measurement endpoint The phenomenon measured in a toxicity test (eg survival, growth or reproduction) that is subsequently related to an assessment endpoint by a risk assessor (Suter, 1993).

Mechanistic model A mathematical model that incorporates scientific knowledge about natural physical, chemical or biological processes (cf **empirical model**).

Monotonic A response that occurs smoothly and unchangingly in one direction.

No Effect Concentration (NEC) The concentration of a material that has no effect on the endpoint measured in a toxicity test.

No Observed Effect Concentration (NOEC) The highest concentration of a material in a toxicity test that has no statistically significant adverse effect on the exposed population of test organisms compared with the controls (Rand, 1995).

Nonparametric A statistical technique that does not assume an underlying distribution for the data (Zar, 1984).

Normal distribution A bell-shaped frequency curve (Zar, 1984, but note that not all bell-shaped curves are normal).

Parametric A statistical technique that assumes an underlying, often normal, distribution for the data (Zar, 1984).

Point estimate A statistical estimate consisting of a single numerical value that summarises a data set. An EC50 is an example of a point estimate (Chapman *et al.*, 1996).

Precision The effectiveness of a measurement process in producing similar results on repeated application (Newman, 1995).

Predicted Environmental Concentration (PEC) The concentration of a material estimated as being likely to occur in environmental waters to which aquatic organisms are exposed as a result of planned manufacture, use and disposal (Rand, 1995).

Probit The 'probability unit' used to linearise responses measured in a toxicity test (Newman, 1995).

Quantal response An all or nothing response such as survival (Newman, 1995).

Regression A statistical technique to determine the functional or predictive relationship between two or more variables (Zar, 1984).

Risk assessment An assessment of the probability of a hazard being realised (Newman, 1995).

Spearman-Karber method A nonparametric statistical technique for analysing quantal data (Newman, 1995).

Stochastic model	A mathematical model that accounts for some of the uncertainty in the output of the system being modelled (Newman, 1995).
Survival time model	A mathematical model that uses time to death as the basis for analysis (Newman, 1995).
Toxic threshold	A concentration above which some effect will be produced and below which it will not (Rand, 1995).
Toxicity curve	The curve obtained either by plotting the median survival times of a population against test material concentrations or median effective concentrations against exposure times (Rand, 1995).
Toxicity test	The means by which the toxicity of a chemical or other test material is determined. A toxicity test is used to measure the degree of response produced by exposure to a specific level of stimulus (Rand, 1995).
Toxstat	A statistical package marketed by West Inc. (1994) that will estimate **LC**, **EC** and **NOEC** values using a variety of parametric and nonparametric methods.
Type I error	The rejection of a null hypothesis when it is true (Zar, 1984).
Type II error	Failure to reject the null hypothesis when it is false (Zar, 1984).
Weibull distribution	A flexible generalisation of the exponential model that may be used to describe and transform individual tolerances in a toxicity test. The flexibility is due to the shape parameter which can give the distribution a positive or negative skew. The Weibull reduces to an exponential model when the shape parameter is 1 (Newman, 1995).

EXECUTIVE SUMMARY

Ecotoxicity tests and bioassays are used to determine the toxic hazard posed by chemicals and environmental samples. The Environment Agency intends to use such tests within a Direct Toxicity Assessment (DTA) programme to set toxicity targets for selected effluent discharges to surface waters. The data from DTA need to be analysed statistically to provide summary measures before they can be used within a risk assessment framework to derive a toxicity target. Several different types of summary statistic are available. This report reviews each of them and describes their advantages and disadvantages. Point estimates of effect, such as the EC50, are preferred over the No Observed Effect Concentration (NOEC). However, standard methods for estimating EC values do not make efficient use of the data. Models that use the time course of toxicity as well as the intensity of effect have many advantages. They can be used to estimate more precise EC values, true No Effect Concentrations (NECs) and chronic effects from acute data sets. Data from pure product testing and the Agency's DTA pilot study are analysed using several different statistical approaches. The results of these analyses are then compared to determine their potential for accurate and precise summaries of toxicity data. The conclusions and recommendations from this report are:

- the NOEC should be phased out as a summary of toxicity in effluent bioassays;
- the Agency should instead use a regression-based estimation procedure alone;
- time should be incorporated in the analytical procedures for DTA data;
- procedures for the collection of data through time should be included in the Agency's Ecotoxicology Methods Guidelines;
- the Agency should determine the optimum spacing of test concentrations, the optimum number of replicates per concentration and the optimal number of organisms per replicate for effluent bioassays;
- predictions of chronic effects from acute effluent bioassay data sets should be tested in the field through the use of macroinvertebrate biomonitoring and *in situ* bioassay techniques;
- the Agency should initiate further study into the survival time models that are available (both mechanistic and empirical) and which of these, if any, best meets the needs of the DTA programme;
- the relationship between different measurement endpoints such as survival, growth and reproduction should be investigated to determine whether there are predictable regularities;
- a steering group should be set up by the Agency to direct the mathematical, statistical and biological work required to take these recommendations forward and to provide a firm underpinning to the quantitative analysis of DTA data.

KEY WORDS

Effluent bioassay, Direct Toxicity Assessment, statistics, NOEC, EC, NEC, survival time

1. INTRODUCTION

The Environment Agency (The Agency) in England and Wales has proposed the application of Direct Toxicity Assessment (DTA) to control selected aqueous effluent discharges to surface waters (Environment Agency, 1996). For the first time in the UK there will be a national strategy for the use of ecotoxicity tests in monitoring and controlling effluents. Ecotoxicity tests are designed to help risk assessors to predict the acute and chronic biological effects of exposure to chemicals. Evidence from the United States suggests that the results from effluent bioassays are useful predictors of toxic effects in surface waters (Dickson *et al.*, 1992; Marcus and McDonald, 1992; Grothe *et al.*, 1996).

Statistical analyses are essential tools for interpreting the outcomes of a bioassay (Cox, 1972; Nelson, 1982; Cox and Oakes, 1984; Crane and Chapman, 1996). This is because bioassays 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 effluent (0 - 100%) 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 reproduction) at each concentration (Forbes, 1993). This response is then used to estimate the concentration-response curve and the point estimates currently required by risk assessors, such as the median lethal concentration (LC50) and the No Observed Effect Concentration (NOEC). Each of these is 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 LC50, while chronic fish tests normally last for several days or weeks and report an Effective Concentration (EC) value and a NOEC. The Agency currently recommends the use of 48-h tests with crustaceans, 72-h tests with algae and 96-h tests with fish amongst a battery of acute effluent bioassays (Environment Agency, 1996; Johnson *et al.*, 1996; Whitehouse *et al.*, 1996).

The response of test organisms to toxicants depends not only upon the dose or concentration to which they are exposed, but also on the duration of that exposure. Risk assessors recognise the importance of this, but formal mathematical methods for incorporating exposure duration currently tend to be eschewed in favour of qualitative graphical approaches (Suter *et al.*, 1987; Suter, 1993, see Figure 1.1), the derivation of acute-to-chronic ratios (Kenaga, 1982), or the use of application factors (Mount and Stephan, 1967). All of the above methods require that *both* acute and chronic tests are performed for at least a proportion of test species and chemicals within an homologous group. This is because an assessor must know what application factor or acute-to-chronic ratio is appropriate for the type of chemical under review, or whether it is likely that a toxic threshold for the substance exists. Such a requirement for chronic testing is unrealistic for many effluents discharged to surface waters. This is because toxicity in these wastes alters over time due to volatilisation, partitioning, hydrolysis and other physical or chemical factors. Hence effluent bioassays are often acute or 'subchronic', lasting for only a few days or hours. This lack of chronic exposure is exacerbated by the limitation of effluent bioassays to a maximum concentration of 100% of whole effluent. Because of this, the 'accelerated' testing of very high concentrations, typical of ecotoxicity testing with industrial and agrochemical products, is not possible, and great uncertainty about chronic effects at the immediate point of discharge remains.

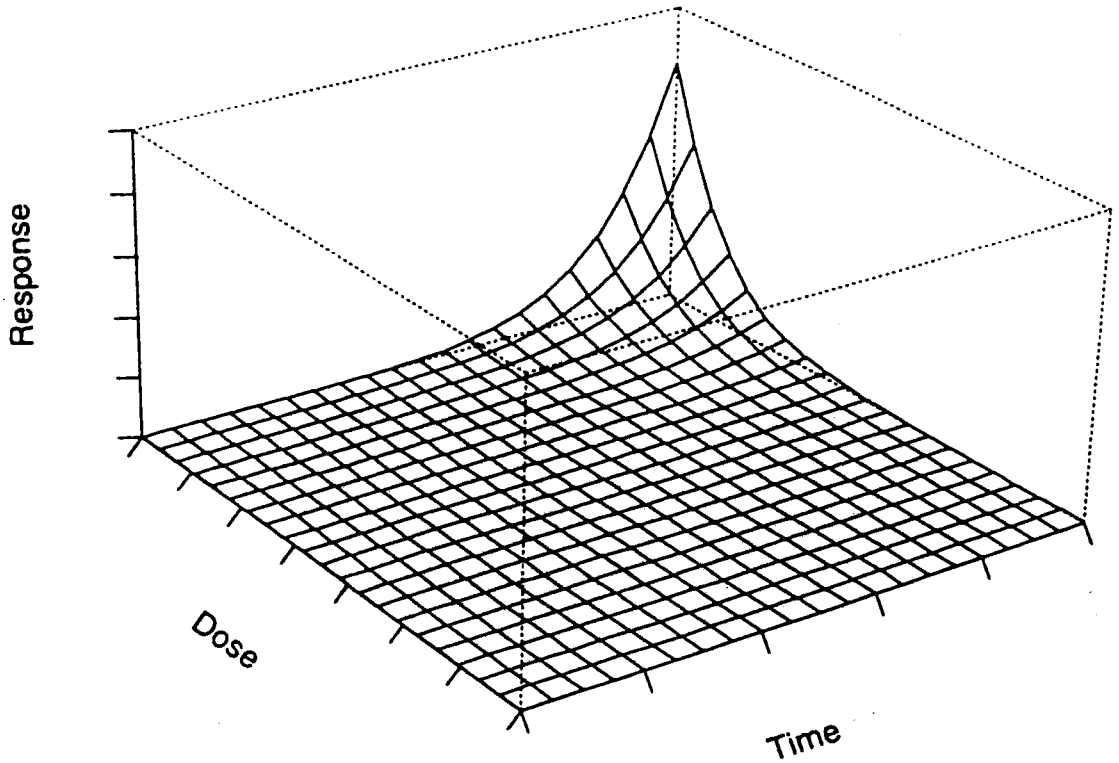


Figure 1.1 Example of graphical approach commonly used by risk assessors.

If the time-course of toxicity could be taken into account it might be possible to make predictions of chronic toxicity on the basis of acute results. This would reduce uncertainty in the derivation of toxicity targets and allow waste dischargers and environmental regulators to reduce effluent toxicity to levels below those likely to cause chronic lethal effects. The incorporation of temporal effects may also help to improve the precision of LC/EC50 estimates and the accuracy of estimates of No Effect Concentrations (NECs). True NECs can be estimated instead of the standard NOEC, which is considered as non-protective by many statisticians (Chapman *et al.*, 1996; OECD, 1996). Newman (1995) also argues that if data on the time course of effects on survival and reproduction are available, ecologically meaningful endpoints, such as the intrinsic rate of population increase can also be estimated.

Several approaches for the estimation of NECs and/or acute:chronic extrapolations have been developed recently. Mayer *et al.* (1991; 1994) have used two-step linear regression and multifactor probit analysis. 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 represent a theoretical improvement on the standard methods and are all discussed later in this report. However, the prediction of chronic lethality and the estimation of NECs are very sensitive to changes in the model on which they are based (Sun *et al.*, 1995). The models should therefore be extensively tested on empirical data before they are used by environmental managers.

This report begins such an evaluation by assessing the theoretical and practical advantages for the DTA programme of currently available statistical summary methods. The summary statistics produced by the different approaches are compared, and recommendations provided on the most efficient, consistent and useful approaches for estimating 'safe' concentrations and chronic effects from acute bioassay data. The specific objectives of this research were,

1. to review available techniques for estimating an NOEC, an ECx, an NEC and a chronic value from acute data;

2. to attend and report on a workshop held by the Organisation for Economic Cooperation and Development (OECD) at Braunschweig, Germany, October 1996 to discuss the statistical options available for analysing the results from toxicity tests performed under OECD guidelines;

3. to analyse data obtained from DTA pilot studies and from other bioassays and toxicity tests to determine the empirical value of different statistical approaches;

4. to recommend optimal statistical methods and identify any further research necessary for meeting Agency needs in this area;

5. to take a view on whether or not the choice of statistical summary affects the value of toxicity targets in improvement plans and to recommend the most pragmatic approach for regulatory testing.

2. REVIEW OF METHODS

Three main approaches to the analysis of ecotoxicity data can be identified (OECD, 1996):

- hypothesis testing to determine an NOEC;
- regression analysis to estimate a time-specific effect concentration such as an LC50; and
- survival time analysis.

Each of these major approaches includes several different specific methodologies. These will be discussed below.

2.1 Hypothesis Testing to Determine an NOEC

The NOEC is the highest concentration in a bioassay producing a response that does not differ from the control when compared in a statistical significance test. The NOEC is usually calculated from concentration-response data by using Analysis of Variance (ANOVA), followed by a multiple comparison test such as Dunnett's or Tukey's test (Newman, 1995). This can only be done if there is replication at each concentration.

The perceived advantage of the NOEC is that it is easy to understand (OECD, 1996). However, there are many disadvantages to its use (Bruce and Versteeg, 1992; Chapman *et al.*, 1996; Crump, 1984; Hoekstra and van Ewijk, 1993; Kooijman, 1981, 1996; Laskowski, 1995; Noppert *et al.*, 1994; Pack, 1993; Skalski, 1981; Stephan and Rogers, 1985). These are:

- The NOEC *must* be one of the concentrations used in a bioassay, because this type of hypothesis testing does not allow interpolation between test concentrations. This means that if concentrations are spaced far apart in an experiment there can be great inaccuracy in determining an NOEC.
- The NOEC tends to increase as the precision of the bioassay decreases, thus rewarding careless experimentation that increases response variability. An example of this is shown in Example 2.1.
- Replication is often low in bioassays, with three replicates per treatment a common choice. This means that the statistical power of the ANOVA and multiple range test is also often low, which also makes it difficult to discriminate between different treatments.
- 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 chosen in significance tests and on the type of multiple range test that is selected for comparing the means: different choices produce different NOECs. In the example provided in Example 2.1 the NOEC ranges between 1% and 46% effluent depending upon which test is used,

- It can be difficult to determine a NOEC if the response does not follow a monotonic trend, for example when stimulation, or 'hormesis', occurs at low concentrations (Figure 2.1).

Such 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 ecotoxicity tests (Chapman *et al.*, 1996), except in the special case of limit tests (Whitehouse *et al.*, 1996).

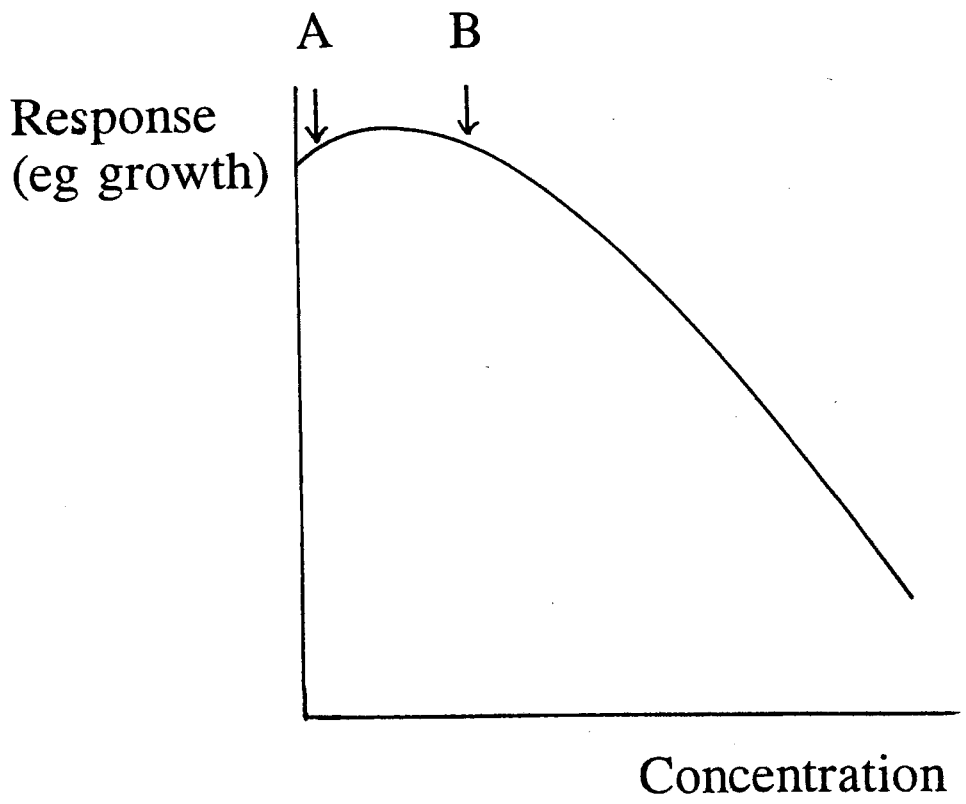


Figure 2.1 There are problems in determining an NOEC when the response curve is not monotonic. In this figure growth is slightly stimulated by low concentrations of effluent. Is the NOEC at point A or point B?

Concentration of effluent (%)	Laboratory 1 results (Survival: 3 replicates)	Laboratory 2 results (Survival: 3 replicates)
0	10, 9, 10	10, 9, 8
1.0	10, 7, 8	8, 6, 10
2.2	9, 6, 7	9, 2, 9
4.6	8, 5, 7	6, 0, 7
10.0	7, 6, 6	1, 7, 6
22.0	5, 4, 2	1, 6, 5
46.0	4, 1, 1	5, 2, 0
60.0	0, 1, 0	0, 0, 0
100.0	0, 0, 0	0, 0, 0
Calculated NOEC (%)		
Dunnett's test	2.2	22
Bonferroni t test	2.2	22
Tukey test	10.0	46
William's test	1.0	2.2
Dunn's test	46	46

Example 2.1 Low precision can increase the NOEC. In this example two effluent bioassays are run with the same effluent at the same concentrations, but in two different laboratories. Laboratory 1 produces more precise results than Laboratory 2. However, the penalty for this is that the NOEC calculated by Dunnett's test for Laboratory 1 is 2.2% and the NOEC for Laboratory 2 is 22%. Calculations were on square root transformed data using Toxstat 3.4 (West Inc, 1994).

2.2 The Estimation of a Time-Specific LC/EC Value

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 affects 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 usually transformed so that least-squares fits can be made to linear models. Linear models have advantages, despite the availability of non-linear curve-fitting routines in most statistical software packages. The estimation of confidence limits is easier and methods for checking model fit are better developed for linear models (Forbes, 1993). Linearity is usually achieved by logging the exposure concentration and converting the response to its probit (Bliss, 1935; Finney, 1971) or logit (Berkson, 1944). Several other linearising transformations are also available, such as the arcsine or Weibull (Weibull, 1951). More recently, Generalised Linear Models have been proposed as a method of analysing data without the need for an initial linearising transformation (Kerr and Meador, 1996).

Whatever the derivation of the concentration-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 usually possible at this median point. However, the x in ECx can be as large or small as an investigator wishes, although estimates at the lower extreme of the probability function are likely to have confidence intervals that pass through zero. It is then impossible to determine whether a chemical has a safe low dose. Hartley and Sielken (1977) summed up this problem as follows: 'Experiments attempting to measure ... minute differential risk increments *directly* by using only extremely small ... *residual doses* are forced, in the face of statistical errors, to use astronomically large numbers of animals. On the other hand, if experiments are conducted at adequately high doses (accelerated doses) the problem of extrapolating (or interpolating) ... to residual dose levels arises.'

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; Figure 2.2). The argument over what x to choose in ECx is probably sterile, as the value of a 'safe' x will depend upon the life history strategy of the organism that is tested. There would be little effect on the population size of herring if 90% of larvae were killed by a toxicant before density-dependent mortality occurred. However, 90% mortality of fish after density-dependent mortality would have a major effect on population size.

Partly because of these problems, much effort has been expended on the estimation of' thresholds' below which no toxic effects occur (Cox, 1987). The problem with thresholds is that they are, like other concentration-response models, highly model dependent. The estimates of thresholds from different models may vary widely, and the data from standard ecotoxicity experiments provide little information on which model is correct (Hoekstra and van Ewijk, 1993; Figure 2.3).

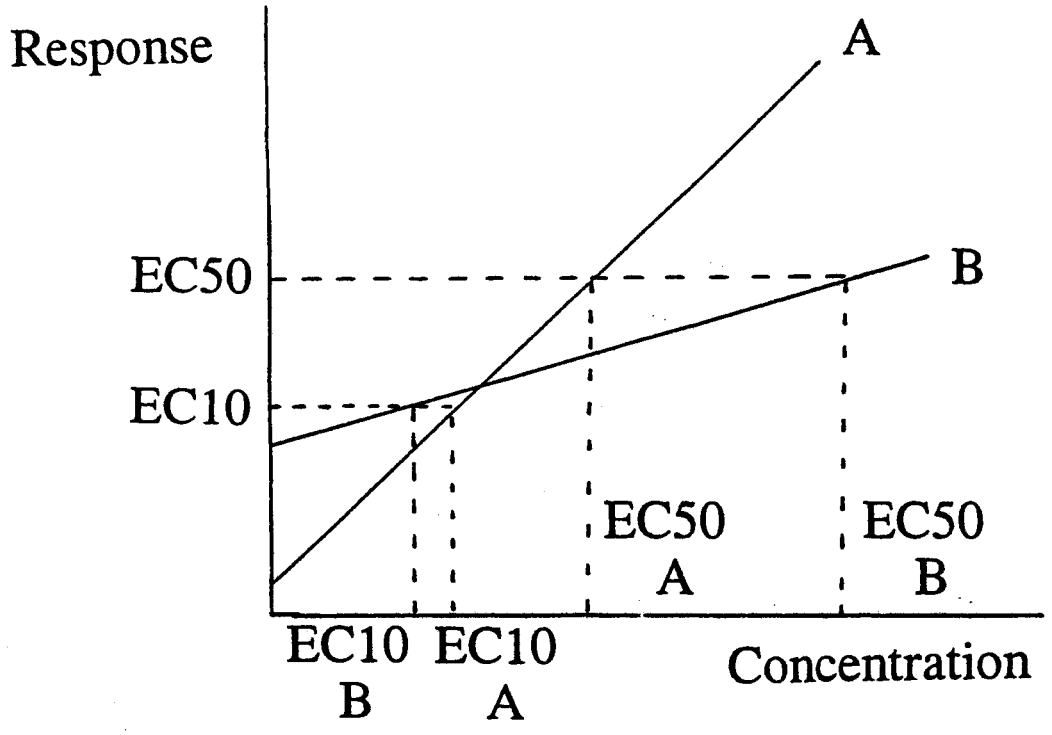


Figure 2.2 **Consistent toxicity ranking depends upon parallel concentration-response curves. In this example the EC50 of effluent A is lower than the EC50 of effluent B, suggesting that effluent A is 'more toxic.' However, the EC10 for effluent A is higher than the EC10 for effluent B, suggesting that effluent A is 'less toxic.' This apparent anomaly can be explained by the difference in the slopes of the two responses.**

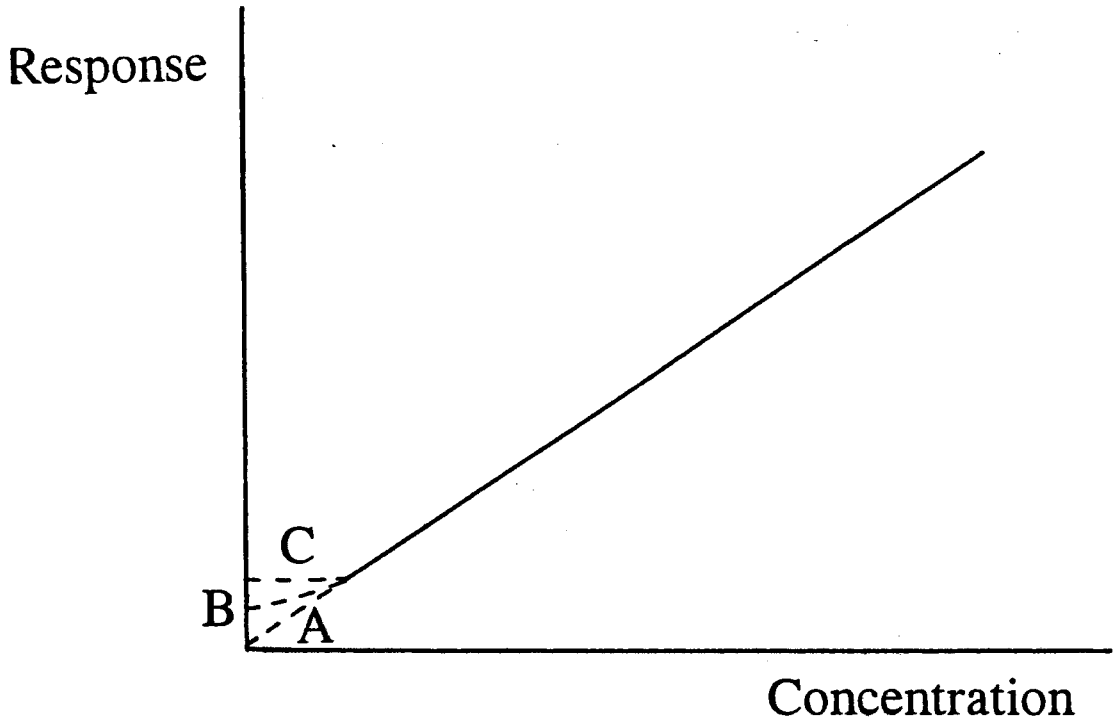


Figure 2.3 The threshold determined in a threshold model is model-dependent. This example shows that there may be no threshold (A) or some empirically unknown threshold that differs depending upon the model used (B and C).

To avoid these apparent shortcomings in the estimation of a 'safe' threshold, Crump (1984) proposed the estimation of 'benchmark' concentrations and Hoekstra and van Ewijk (1993) proposed 'bounded effect' concentrations. These are approaches that do not rely so heavily on model assumptions.

Benchmark concentrations, as defined by Crump (1984), are the lower limit of the 95% confidence interval estimated for a specific effect in the range of 1-10%. The 1-10% effect is estimated by model-based extrapolation, so problems of model dependency remain. However, Crump (1984) argues that the differences between different model predictions are relatively small in this range (Figure 2.4). This is not the case for extrapolation below this range, and efforts to do this (eg Chen and Kodell, 1989) have been criticised as flawed (Hoekstra and van Ewijk, 1993).

Bounded-effect concentrations (Hoekstra and van Ewijk, 1993) are estimated by first selecting a concentration (the bounded-effect concentration) for linear extrapolation to a lower concentration. Hoekstra and van Ewijk (1993) suggest that the concentration corresponding to $\leq$25% effect is often a justifiable choice for the bounded effect concentration. If 25% effect is chosen as the criterion, then the bounded effect concentration should be the highest *tested* concentration for which the upper confidence limit does not exceed 25% effect. Next, the concentration at which an acceptably small effect occurs (eg 1% effect) is estimated by linear extrapolation from the bounded effect concentration. For example, if in an effluent bioassay a concentration of 10% effluent produces an effect with an upper confidence limit of 24%, the 1% effect would be estimated by extrapolating from the 10% bounded effect concentration:

$x/10 = 1/24$

$x = 0.42\%$

An effluent concentration of 0.42% would then be considered a 'safe' concentration. However, it is important to note that this 'safe' concentration will apply only to an exposure duration similar to that in the experiment. Benchmark and bounded-effects methods do not attempt to predict a chronic NEC from the time-course of toxicity. This is a problem with all time-specific summaries of toxicity test data and is one reason why extrapolation or safety factors are applied. Reliance on safety factors may be reduced if the time course of toxicity is taken into account.

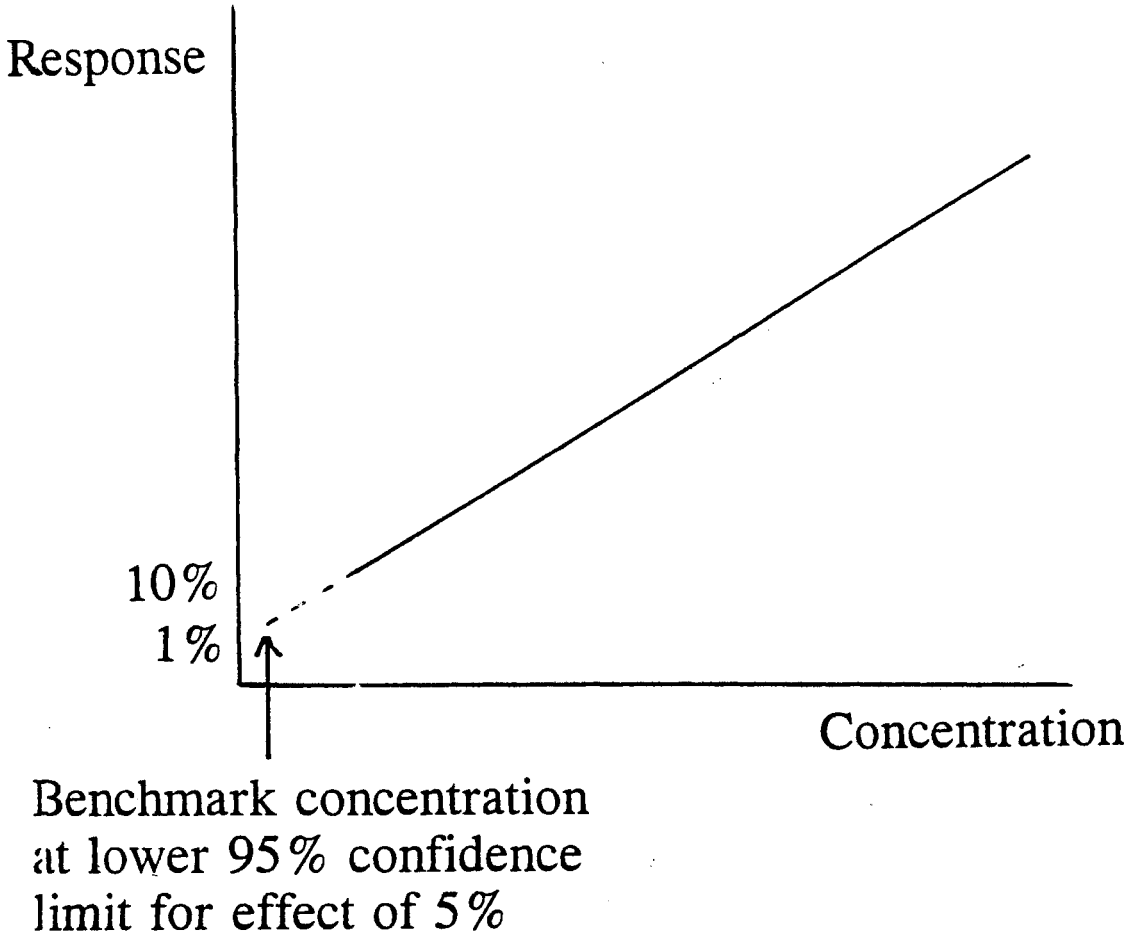


Figure 2.4 Example of a benchmark concentration (Crump, 1984).

2.3 Analyses Including Exposure Duration

Although there is general agreement that use of a concentration-response curve to estimate an ECx or a threshold, benchmark or bounded effect concentration has many advantages over the derivation of a NOEC, the calculation of an ECx at specific time intervals still does not use all of the available data. 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.

2.3.1 Standard analyses of exposure duration

The importance of the duration of exposure in aquatic toxicity tests has been recognised for many years (eg Powers, 1917). Bliss and Cattell (1943) suggested in an early paper that quantal assays based upon the reaction time of individuals were more efficient than those based upon measurement of the distribution of thresholds within a population (the standard approach discussed in Section 2.2). Traditionally, the standard analysis of exposure duration was to plot the mean or median survival time (or its reciprocal) against the toxicant concentration (or log concentration) (Abram, 1964, 1967; Alabaster and Abram, 1965; Alderdice and Brett, 1957; Gaddum, 1953; Herbert and Shurben, 1964; Lloyd, 1960; Sprague, 1969). A line was then fitted either by eye, or by using a more formal model such as the rectangular hyperbola (Hey and Hey, 1960). A threshold level of chemical or effluent over an indefinite period was then estimated from this (eg Alderdice and Brett, 1957; Figure 2.5).An alternative is to plot the toxicity curve of the median lethal concentration against time (eg Heming *et al.*, 1989; Figure 2.6).

The problem with these approaches is that they may involve lengthy experiments and, for some chemicals, biological effects may not even then reach an asymptote within the duration of the experiment. Because of these problems several workers have attempted to use all of the data from much shorter-term experiments to produce more precise estimates and predict potential chronic effects.

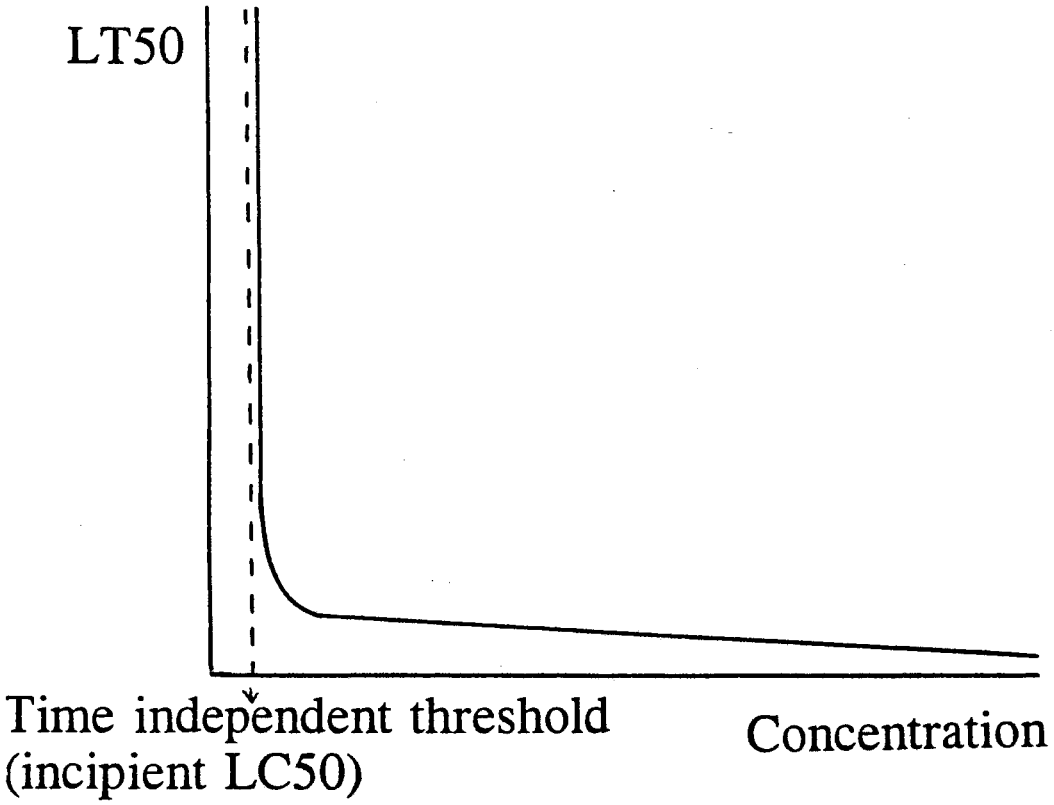


Figure 2.5 **Standard analysis of toxicity over time**

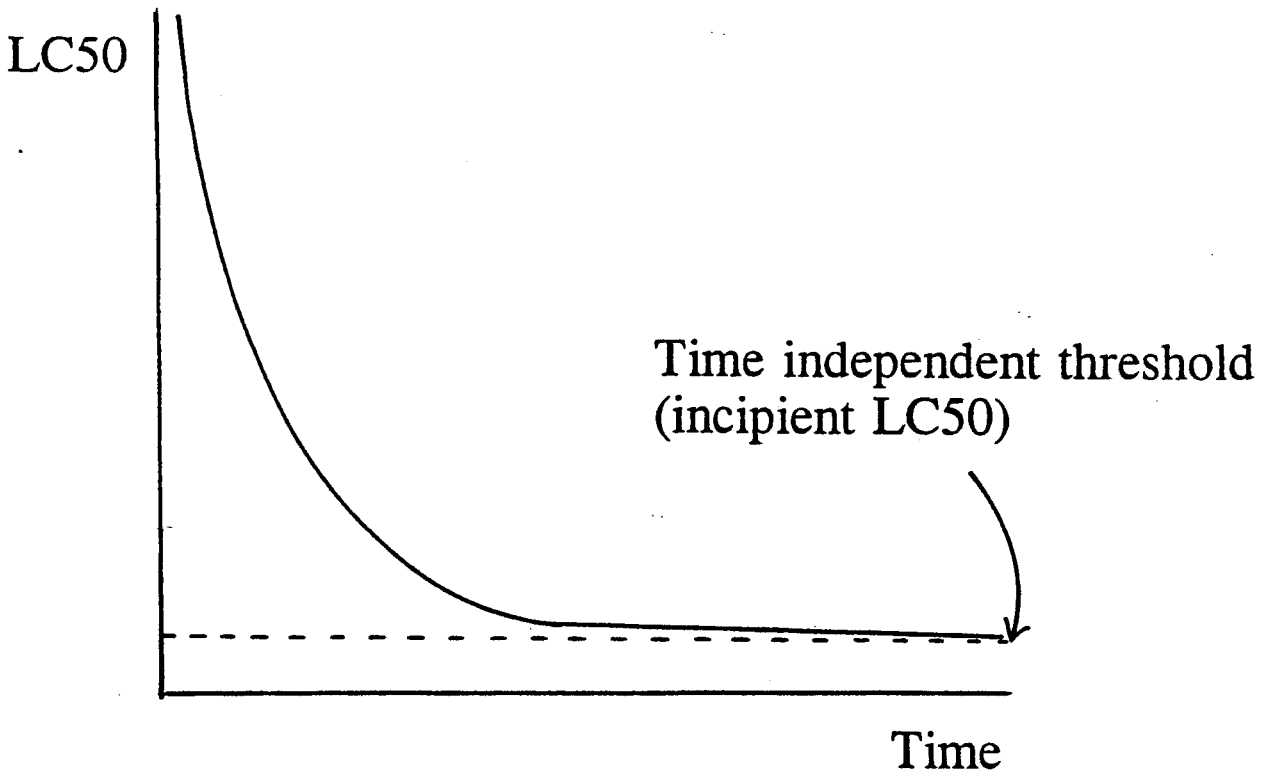


Figure 2.6 **Example of a toxicity curve**

2.3.2 Recent advances in the analysis of exposure duration

Two-Step Linear Regression and Multifactor Probit Analysis

Mayer *et al.* (1991, 1994) developed a simple statistical approach for using observations at different time periods in acute tests to predict chronic lethality. They made two assumptions essential for any prediction of this type, i) the concentration response is a continuum in time, and ii) the mode of action and detoxification systems for lethality are similar under acute and chronic exposures. Linear regression is used to estimate an LC0 at all observation times (probit percentage mortality = a + b(log concentration)). These estimates of LC0 are then regressed against the reciprocal of time (LC0 = a+b(1/t)). The intercept of the regression line is the chronic 'predicted no observed effect concentration' (PNOEC; Figure 2.7).

Using the reciprocal of time may be sufficient to produce a straight line from the usual hyperbolic relationship between LC values and time (Green, 1965), but Mayer *et al.* (1991, 1994) found that additional log transformations (log LC0 = a + b(1/t) or log LC0 = a + b (1/log t)) were occasionally necessary because of negative intercepts or curvilinear data (Mayer *et al.* 1994). The disadvantage of this approach is that the estimation of confidence limits is uncertain due to the additive nature of variation in the two-step procedure. Kooijman (1981) has rightly advised caution when extrapolating effects beyond those that are observed, and strongly recommended the use of confidence intervals rather than point estimates alone.

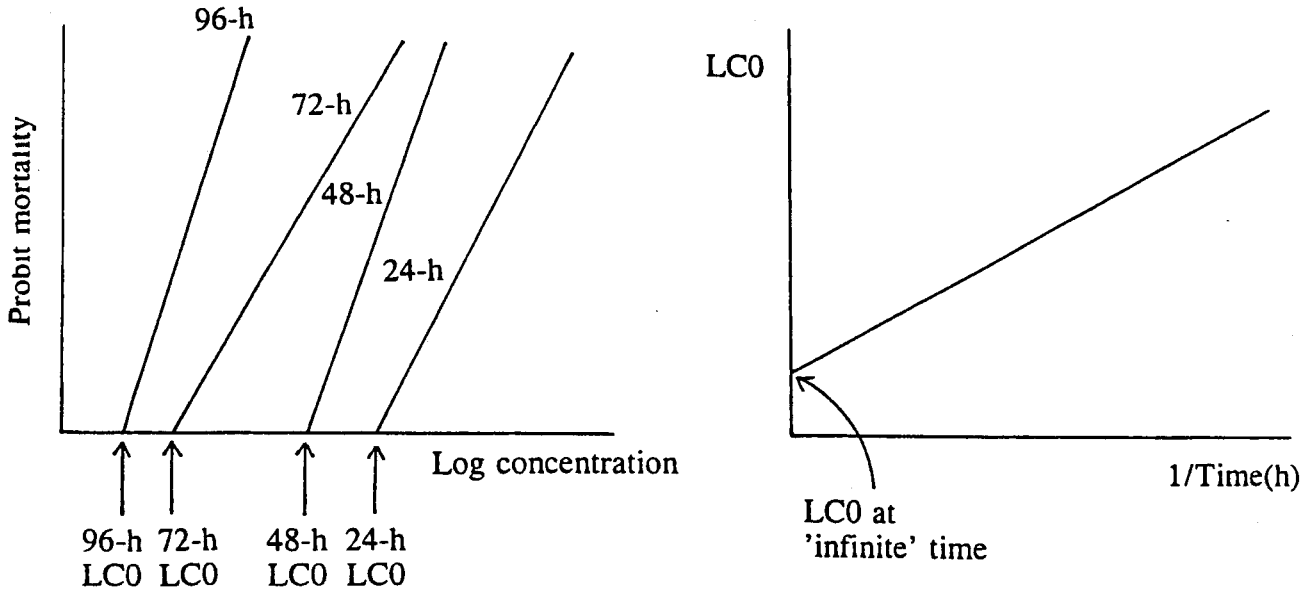


Figure 2.7 Two-step linear regression

Where the test data are suitable, multifactor probit analysis (MPA) may be used, rather than the two-step linear regression described above (Lee *et al.*, 1995; Mayer *et al.*, 1991, 1994). Data for this method should ideally consist of at least three and ideally five concentrations with mortalities >10% and <90% over a fixed time, and a minimum of four observation times. Mayer *et al.* (1996) recommend that MPA is most appropriate when different experimental units are assessed at different concentration-time combinations. For example, different tanks of fish

would have to be observed at 24, 48, 72 and 96 hours at each test concentration. This is not the normal design for an ecotoxicity experiment, in which the same tanks are observed at all time periods, and would add substantially to costs. It is therefore difficult to forsee a situation in which the use of MPA would be justified.

The two methods described above were used to predict chronic toxicity from acute data for several different fish species and test chemicals. These predictions were then compared with empirically-derived data in the literature and found to agree very well (Mayer *et al.*, 1991, 1994). The authors concluded that chronic toxicity tests with fish were not necessary if lethality, and probably growth (Mayer *et al.*, 1986), were the endpoints of interest, because there appeared to be predictable correlations between survival and growth. However, effects on reproduction could not be predicted accurately from survival and growth data (Suter *et al.*, 1987). Mayer *et al.* (1991) also recommended that the application of the approach to invertebrates required further investigation. Later in this report we use the same data set as Mayer *et al.* (1991) for comparing the results of different techniques.

Survival time modelling

Recently, Sun *et al.* (1995), in a continuation of the work described above, have recommended the use of survival time modelling and accelerated life testing. This has an advantage over the earlier work by Mayer *et al.* (1991, 1994), in that the statistical dependence of observations at different times does not present analytical difficulties. Survival time models have also been proposed by other researchers as a method for integrating time, concentration, response, and ecologically important covariables such as organism weight and sex (Diamond *et al.*, 1989; Dixon and Newman, 1991; Newman and Aplin, 1992; Newman *et al.*, 1989, 1994). These models belong to the family of accelerated failure time models, commonly used by engineers and medical statisticians (Kalbfleisch *et al.*, 1983). Sun *et al.* (1995) have described a model in which organism survival at a particular period of time in a particular concentration (Q(t,x)) depends upon the strength of toxic action (a), the concentration-response surface shape (b) and the time-response surface shape (c). The mathematical expression of this is $Q(t,x) = \exp\{-ax^b t^c\}$.

This particular formulation is based upon a Weibull distribution, but other statistical distributions could be used, depending upon prior knowledge. For example, Newman *et al.* (1994) made their choice from the normal, log normal, Weibull, gamma and log-logistic distributions by comparing the fit of each of these to the data. However, discriminating between the fit of the typical models used in toxicity testing can be difficult without prohibitively large sample sizes (Prentice, 1976), and Newman *et al.* (1994) found that several models had very similar fits. User-friendly software has been produced by Sun *et al.* (1995) so that predictions of chronic toxicity using this approach can be made by non-statisticians. The software enables the user to specify a particular level of effect that is 'acceptable' (eg 1%, 0.1%, 0.01%, or 0%) and to estimate the chemical concentration likely to produce such an effect. Newman (1995) prefers to use the survival time approach to produce estimates of ecologically important population parameters, such as an increase in relative mortality risks, or changes in the intrinsic rate of population increase. The choice of statistical summary by a risk assessor will depend upon the legislation currently in force and upon their specific choice of assessment endpoint (Suter, 1993).

The survival time approach outlined above is almost entirely empirical. Statistical techniques are used to fit models to the data collected in bioassays and the use of biological theory is restricted to the choice of an appropriate distribution. Some survival time approaches do not even require

assumptions about distributions. The Kaplan-Meier approach (Kaplan and Meier, 1958) is nonparametric and implemented in many user-friendly software packages such as Unistat (Unistat Ltd., 1994). These empirical survival time approaches therefore have the advantage that subjective assumptions in the models are rather few, but the disadvantage that biological information that might be used to reduce error estimates remains unused.

Kinetic models

In contrast with the empirical models described above, there are some approaches that use theory from biology and chemistry within models that incorporate toxicant concentration and exposure duration. For example, several authors have proposed the use of simple kinetic models to overcome the problem of determining NEC values, ECx at infinite exposure time, or both of these (eg Chen and Selleck, 1969; Chew and Hamilton, 1985; Heming *et al.*, 1989; Matida, 1960). Compared with the empirical models described above, these kinetic models involve more biological assumptions about the behaviour of toxicants in living organisms (Kalbfleisch *et al.*, 1983). The method proposed by Matida (1960) is one of the simplest, and involves fitting a one-compartment first-order kinetics model to LC50 and duration data in a manner very similar to the derivation of a toxicity curve in Section 2.3.1. Although this approach is not recommended statistically, because the LC50 data are not independent, it can be useful if the analyst is presented with only LC50 data, rather than the complete set of raw data (Rand *et al.*, 1995). However, it is preferable to use a more sophisticated and valid statistical method if time-to-death data are available.

For example, Chen and Selleck (1969) suggested that when organisms are exposed to toxic chemicals there is an induction period (t_i) in which no effects are observed, because toxicity is unlikely to be spontaneous. The formal mathematical expression of this is $dN/dt = 0$; $0<t<t_i$. N in this equation is the number of organisms surviving at a particular exposure time and t is a rate coefficient. When t is greater than t_i the concentration of the toxicant in the organism can be expressed by the equation $dN/dt = - KC^nN + HN$; $t > t_i$. In this equation, K and H are also rate coefficients and n is the order of reaction. These parameters are all derived by regression analysis from a concentration-response test. The term $- KC^nN$ is the rate of toxification and the term HN is the rate of detoxification (Figure 2.8). These simple equations can be rearranged and integrated to provide valuable summary statistics, such as the threshold concentration (C_t) and the survival ratio (N/N_0).

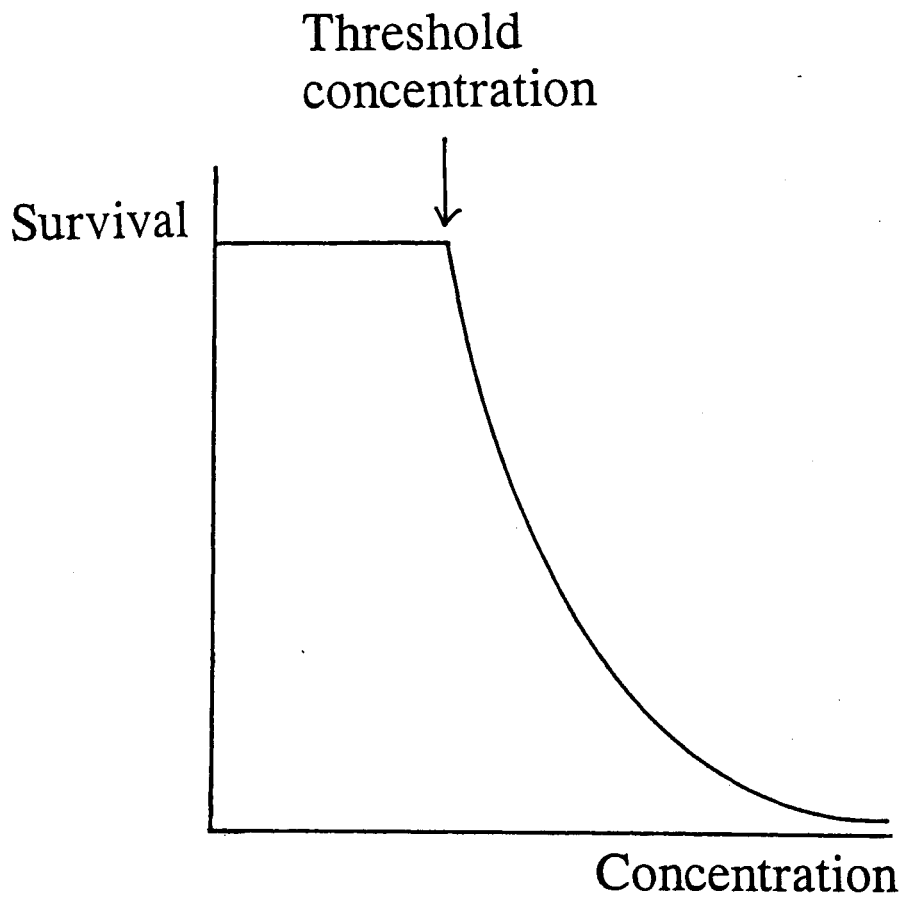


Figure 2.8 Chen and Selleck's (1969) kinetic model.

Kooijman (1981, 1987) used ideas similar to those of Chen and Selleck (1969) in developing a model to describe the relationship between toxicity and time and was also able to produce a model of the relationship between acute and chronic LC50 values. Recently, however, Kooijman has rejected his earlier attempts to 'improve' standard models for analysing toxicity data (Kooijman, 1996). He points out that standard analyses often treat toxic effects as deterministic at the level of the organism, but stochastic at the population level. In other words, individuals die when a fixed toxic threshold is exceeded, but that threshold is a random trial from a statistical distribution for the population. Kooijman (1993, 1996) believes that there is now sufficient theoretical and empirical evidence to suggest that effects at the individual level are also stochastic. This then allows the use of hazard-based models that replace estimates of the LCx and the slope of the response with a single parameter, the killing rate (defined as the probability of dying, per unit of time and per unit of environmental concentration that exceeds the NEC (Bedaux and Kooijman, 1994)). Hence, the new model involves the estimation of only three parameters (NEC, killing rate and elimination rate), compared with the four required in Kooijman's (1981) model (LC50, slope of response, NEC and elimination rate).

The hazard rate of a substance is estimated from the following,

$$h(t,c) = k_1(c(1-e^{-ket}) - c_0)_+$$

where

$h(t,c)$	=	hazard rate (the probability of dying during a given exposure period)
t	=	duration of exposure
c	=	concentration of substance in environment
k_1	=	killing rate
k_e	=	elimination rate of the substance
c_0	=	NEC of substance in environment
$(x)_+$	=	maximum of x and 0

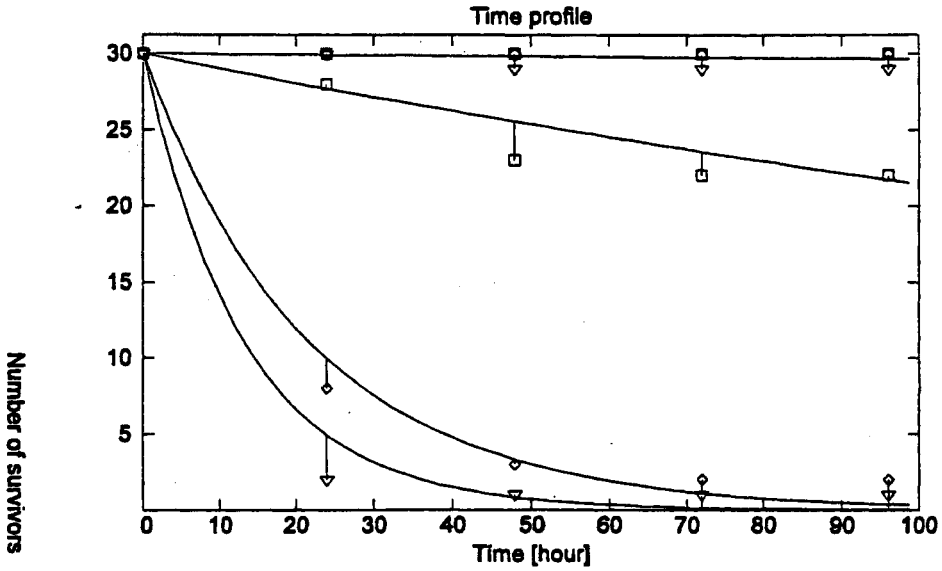
A comparison of this model with the earlier extension of the standard model (Kooijman, 1981) showed that the prediction of effects after prolonged exposure was more realistic with the hazard model (Kooijman, 1993; Bedaux and Kooijman, 1994). Simulations also show that the hazard model is more likely than the extended standard model to produce a positive, non-zero estimate of the NEC, which is of obvious practical advantage (Kooijman and Bedaux, 1996d). The mechanistic basis of this model allows the estimation of toxic effects on reproduction and growth in different species using the same approach (Kooijman and Bedaux 1996a, b, c; Kooijman *et al.* 1996). A further advantage is that it is possible to model the effects of episodic exposure to toxicants and the effect of non-persistent chemicals that degrade relatively rapidly (Widianarko and van Straalen, 1996). However, a potential problem with the hazard model is that accurate estimation of k_1 may require fairly large sample sizes (Bedaux and Kooijman, 1994). Another problem is that the model includes a number of assumptions that may not be transparent to a naive user. For example, one of the assumptions is that effects are exponential, but Newman (1995) provides an example in which the exponential model fitted survival data less well than several others. Kooijman and Bedaux (1996d) themselves recommend that the goodness of fit of the model should always be inspected graphically to ensure a common sense interpretation, but this requires some basic knowledge of statistics that most ecotoxicologists unfortunately appear to lack. A software package, Debtox, is available for making the calculations and producing the graphs described above and is one of those used in the next section of this report (Kooijman and Bedaux, 1996d; Figure 2.9)

Code : 0001
Experiment : Survival, Model : Hazard
Compound : Butyl benzyl pthalate
Species : Fathead minnows
Experimentalist : Monsanto

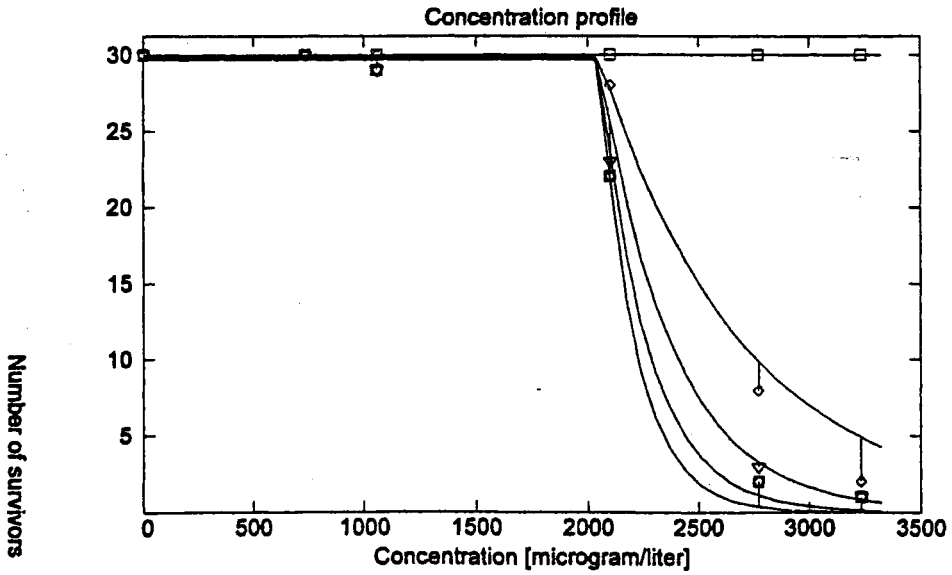
Measurements: Number of survivors

Time	Concentration: microgram/liter					
hour	0	740	1060	2100	2770	3230
0	30	30	30	30	30	30
24	30	30	30	28	8	2
48	30	30	29	23	3	1
72	30	30	29	22	2	1
96	30	30	29	22	2	1

Hazard model, Fast kinetics			ASD	Correlation coefficients	
Blank mortality rate	0.0001165	h^{-1}	0.000		
No-effect concentration	2049	$\mu g\ l^{-1}$	20.948	0.087	
Killing rate	6.375e-005	$1\ \mu g^{-1}h^{-1}$	0.000	0.000	0.495
Deviance	29.52				



(note: the lines in this figure are estimates of survival through time at different concentrations)



(note : the lines in this figure are estimates of survival in specific concentrations at different bservation times)

Figure 2.9 Example output from Debtox (Kooijman and Bedaux, 1996d)

3. EMPIRICAL TESTING OF METHODS

Section 2 above describes the development of data analysis techniques in ecotoxicology from early graphical to computer-intensive approaches. Currently, the main summary statistics from a toxicity test or bioassay are an ECx or an NOEC. In this section we use data generated from two main sources and apply several different statistical techniques to arrive at quantitative summaries of the data.

3.1 Sources of Data

The two main sources of data used in this report were,

- Mayer *et al.*'s (1991) data set for fish mortality in pure chemical and effluent tests. The data set consists of observations at different times on mortality during acute tests, mostly lasting for 96-h. Mayer *et al.* (1991) also report the results of chronic tests with the same chemicals. This allows us to compare model based predictions of chronic toxicity with empirically based measurements;

- Effluent bioassay results provided by the Agency from a series of pilot studies (Environment Agency, 1997). *Daphnia magna* immobilisation over 48-h was the most sensitive higher organism test for four effluents discharged to freshwaters. These data were also amongst the few from the pilot studies that included observations at different time periods (0, 24 and 48-h). The effluents were discharged by three sewage treatment works with industrial inputs and by a chemicals manufacturer. Thirty-six separate *Daphnia* tests were performed on these effluents, of which 30 are analysed below. Data from the remaining six bioassays were not used because there were no mortalities at the highest concentration tested.

Because the data from the pilot studies described above were rather poorly behaved when analysed, a further effluent bioassay dataset was selected. This consisted of 96-h test data for brown shrimp *Crangon crangon* provided by Mike Mallett of Eurolaboratories Ltd. This species has not been selected by the Agency for routine use in the DTA programme because it is less sensitive to toxicants than many other species (D. Forrow, Environment Agency, pers. comm.). However, Anglian Region of the Environment Agency have used it for several years when testing effluents discharged to coastal waters. Five brown shrimp data sets were selected at random for analysis, simply to determine whether there was any intrinsic problem with effluent bioassays that prevented the use of the methods described above.

3.2 Analytical Techniques

The statistical techniques that were used to analyse the data were,

- Estimation of a time-specific ECx. Toxstat version 3.4 (West Inc., 1994) was used to estimate an LC50, LC20 and LC10 and associated 95% confidence intervals (CI) for each data set, where the data quality allowed. Probit (parametric) and Spearman-Karber (nonparametric) models were used to produce estimates, where possible. Neither probit nor Spearman-Karber models are appropriate if there are no partial mortalities but only

total survival and total mortality. In datasets where this was the case, the binomial method (Stephan, 1977) was used to derive an LC50 and 95% CI.

- Estimation of an LCx from both exposure and duration data. Debtox version 1.3 (Kooijman and Bedaux, 1996d) was used to estimate an LC50, LC20, LC10 and LC1 at the end of each test, using data from all observation periods during the tests.

- Estimation of an NEC. Debtox version 1.3 (Kooijman and Bedaux, 1996d) was used to estimate a No Effect Concentration and 95% confidence interval, using data from all observation periods during the tests.

- Prediction of chronic effects from acute data. The Two Step Linear Regression approach and the Accelerated Life Testing approach were used in the Acute to Chronic Estimation software package (ACE, Mayer *et al.*, 1996) to estimate LCx values at infinite time.

The ratios between pairs of data sets were compared and the mean, standard deviation (SD) and range of these ratios calculated. This was done to determine the predictive accuracy of each technique. For example, a ratio of 1 between a predictive method and a measured chronic NOEC shows that the prediction was very accurate. A ratio of 0.1 shows that the prediction overestimated the NOEC by an order of magnitude and a ratio of 10 shows that the prediction underestimated the NOEC by an order of magnitude.

The fish data from Mayer *et al.* (1991) were then logged, and the DTA *Daphnia* data arcsine transformed before further analysis to normalise the distributions (Zar, 1984). Ordinary least squares linear regression was used to analyse pairs of parameters and derive a regression equation, the significance of the intercept and slope, and the coefficient of determination (r^2). Pearson's correlation coefficient (r) was also calculated. These summaries can be used to determine whether there is a linear relationship or a correlation between the x and y values and how strong this relationship might be.

All calculations were performed using Unistat Version 3.0 statistical software (Unistat Ltd., 1994).

3.3 Results And Discussion

3.3.1 Fish Toxicity Data Set

Table 3.1 shows comparisons between summary statistics for the fish toxicity data published by Mayer *et al.* (1992). The first column identifies the datasets that were compared and whether they were the x or the y variable in subsequent comparisons. The next three columns give information on the ratios between the x and y values. The first of these provides a mean ratio for all pairs of data that were compared, the second gives a standard deviation (SD) for this mean and the third shows the range of ratios in the data set. The next eight columns give information from a linear regression performed with the data. The first two provide the regression coefficients a and b, so that the reader can reconstruct the relationship by substituting the coefficients into the usual linear equation $y = a + bx$. The next column gives the degrees of freedom (df) for the regression. The two columns headed 'intercept' give the t statistic and p value for the intercept, which show whether the regression line intercepts at a point other than zero. The two columns headed 'slope' provide the same information for the

slope of the line and allow the reader to determine whether there was a significant positive or negative relationship between the two datasets. The r^2 value is the coefficient of determination, which is a measure of how well the data fit the linear model (1 = perfect fit). The r value is the correlation coefficient, which is simply another way of expressing the strength of the relationship between the two datasets. The data are presented in this way, rather than graphical form (see example in Figure 3.1) since graphs of these data would have looked very similar, because of the need to log the variables.

Debtox estimates of the NEC, EC10 and EC1 generally underestimated the empirically determined NOEC by an average factor of 4 or 5. However, the lower 95% CL of the Debtox NEC estimate tended to overestimate the chronic NOEC by a factor of about 1.5. The predictive power of the latter relationship appears particularly high for the transformed data, with an r value of 0.96, although this tends to disguise some major differences. For example, the chronic NOEC of EPN to sheepshead minnows was 4.1 μg/L (Lowest Observed Effect Concentration (LOEC) 7.9 μg/L), but Debtox predicted a lower 95% CL for the NEC of 124.3 μg/L. In contrast, the chronic NOEC of methyoxychlor to brook trout was 1.1 μg/L (LOEC 3.1 μg/L), but Debtox predicted a lower 95% CL for the NEC of 0.0372 μg/L.

ACE two-step linear regression produced more accurate estimates of the chronic NOEC than Debtox and the range of ratios between ACE estimates and the chronic NOEC was narrower. The maximum underestimate of the chronic NOEC was by a factor of 12 and the maximum overestimate was by a factor of 7. The ACE output produces different model results and the user is advised to select the result with the highest r^2 value. This can have the paradoxical effect, as shown in Table 3.1, of chronic EC10 predictions overestimating the true chronic NOEC, EC1 predictions underestimating the NOEC, and EC0.01 again overestimating the NOEC! However, the final results suggest that the technique is empirically useful, despite this anomaly, and that chronic EC0.01 predictions have high predictive power. Anomalies also occurred when Debtox was used. For example, if the data were poor, the estimate of the 95% CI could be zero and the estimate of the NEC might be greater than the EC50 estimate.

Four of the ACE EC0.01 predictions from the 28 data sets (14%) produced a negative lower 95% CL, thus leaving uncertainty with regard to the existence of a safe level of chemical. This never occurred for this dataset with the Debtox estimates. As would be expected from the above, ACE estimates of the EC1 and EC0.01 tended to be lower than Debtox estimates of the NEC by an average factor of 3.6 and 7.5 respectively. Accelerated Life Testing (ALT) results were generally very poor predictors of chronic NOEC values. The results are not reported in Table 3.1 because they were so poor that we suspect a software problem. Once this problem has been resolved we recommend further testing of the ALT models, as Mayer *et al.* (1996) reported that they gave good predictive power.

The use of time-specific probit, Spearman Karber or binomial methods to estimate 96-h EC50, EC20 and EC10 values always overestimated the chronic NOEC, on average by a factor of 3.2 - 4.6. However, there was close agreement between the 96-h EC values estimated by approaches that were time-specific or incorporated both time and concentration. This agreement was particularly close when Toxstat 96-h EC50 estimates were compared with Debtox 96-h EC50 measurements. The two approaches produced estimates with greater differences for smaller EC values, but there was still no more than a threefold difference.

3.3.2 DTA Pilot Study Data Set

Unfortunately, ACE two-step linear regression could not be used to analyse the *Daphnia* data from the DTA pilot study. The programme was unable to use the limited information provided within this data set to perform the probit analyses that are a necessary first step with this approach.

Table 3.2 shows comparisons between the summary statistics that could be calculated. These comparisons are set out in the same way as in Table 3.1. Debtox estimates of the NEC, EC10 and EC1 consistently suggested that the acute NOEC, currently proposed for setting a toxicity target, was unprotective. The acute NOEC was on average 6.7 times higher than the predicted NEC and 440 times higher than the lower 95% CL of the NEC. In contrast, the Debtox 96-h EC10 and EC1 estimates were in agreement with the acute NOEC and the Toxstat 96-h EC estimates.

Since the Debtox predictions of chronic effects in the analysis of Mayer *et al.*'s (1992) data appear on average to be *underprotective*, it is of some concern that the Debtox estimates of chronic toxicity for the pilot study data suggest that the current approach for deriving a toxicity target is also underprotective. This means that there may still be very high chronic toxicity in surface waters receiving toxicity controlled effluents. A further concern is that the quality of the data from the pilot study was generally insufficient for adequate statistical analysis, even though one of the 'best-behaved' data sets was selected for use in this report. Many of the *Daphnia* data comprised all or nothing responses, with one concentration producing no effect and the next highest concentration producing total mortality. With concentrations spaced relatively far apart, this produces inaccurate and imprecise estimates of summary statistics. Evidence for this is that 22 of 30 Debtox estimates of the NEC lower 95% CL (73%) were zero.

3.3.3 Brown Shrimp Bioassay Data

The difficulty in adequately analysing the DTA pilot study data was not due to inherent difficulties in effluent bioassays. Five sets of results from the brown shrimp bioassay data were successfully analysed using all of the techniques, including ACE two-step linear regression. These analyses suggest that the reasons for the difficulties encountered in analysing the DTA pilot study data were the lack of observations at several time intervals and the lack of partial effects in many of the pilot study data. This latter problem can be addressed either by choosing more appropriate concentrations or by making more frequent observations.

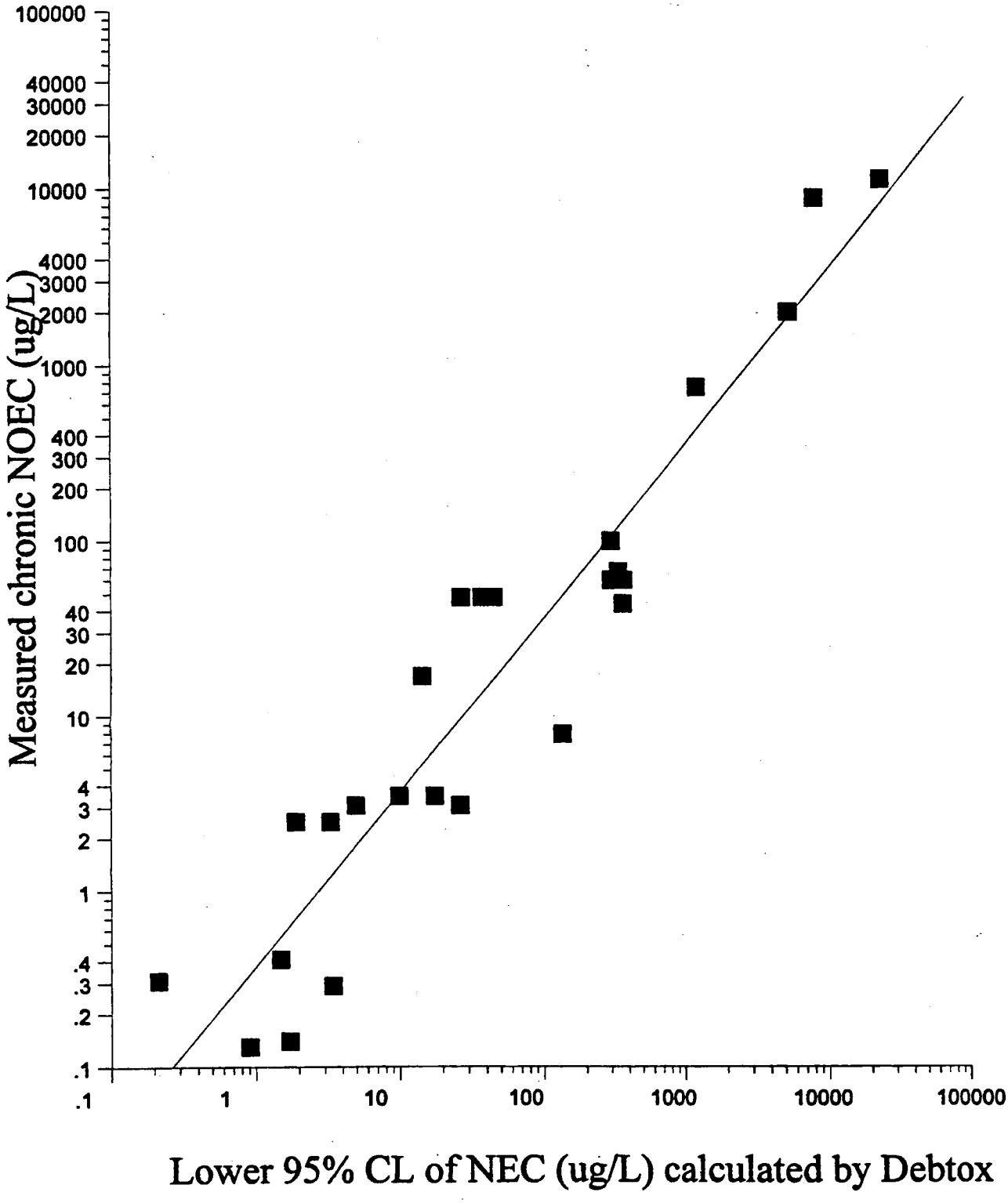


Figure 3.1 **Example of graphical comparison between variables in this section of the report.**

4. GENERAL DISCUSSION

There is a wide range of techniques available for analysing toxicity data, but environmental toxicologists do not appear to be making full use of them. The NOEC is considered by many to be a summary statistic with a very low information content (Chapman *et al.*, 1996). However, even time-specific EC estimation does not use all of the data gathered in most ecotoxicity tests (Newman, 1995). It is because of these concerns that the Organisation for Economic Cooperation and Development organised a meeting of experts in Braunschweig, Germany in 1996 (OECD, 1996). The report from this meeting is added as an appendix to this report.

The main conclusions from the Braunschweig meeting were that bioassays should provide information that is accurate, precise, and interpretable to the non-expert and that information on the time course of effects should be integrated with information on the concentration-response curve. However, it was acknowledged that the specific type of information required from a test will largely depend on the way in which it is extrapolated to the natural world. Ideally,

- the danger of false negatives (Type II errors) should be reduced;
- tests should focus on biologically significant endpoints;
- it should be possible to link bioassay results to predictive ecological and biological models and incorporate biologically important covariates, where this is appropriate;
- better use should be made of all test measurements;
- methods guidelines should make explicit recommendations on appropriate statistical analytical techniques;
- summary statistics and parameters should be simply interpretable by non-statisticians and decisionmakers.

Would the use of the different models examined in this report and recommended to the OECD influence the way in which the Agency derives and monitors toxicity based targets? Currently, the Agency proposes to set toxicity targets by comparing the acute NOEC derived in an effluent bioassay with the Predicted Environmental Concentration (PEC) of the effluent in the surface water to which it is discharged. If the available dilution in this surface water is insufficient to bring the concentration of effluent down to a level below the acute NOEC then toxicity reduction will be required (Johnson *et al.*, 1996). If there is sufficient dilution then no toxicity reduction will be required. Clearly, if chronic toxicity is predicted to occur at effluent concentrations several orders of magnitude lower than acute toxicity there may be many more dischargers who would be required to reduce their effluent toxicity. This could potentially increase the number of toxicity-regulated discharges quite considerably.

The results presented in this report suggest that time of exposure can and probably should be used in the analysis of effluent bioassay data. However, many different models are available, each with their own assumptions. The Agency needs criteria for appropriate selection from the available mechanistic and empirical models. The best reasonable fit should be the most important criterion for model selection: models should be as simple as possible and only as complex as necessary. Mechanistic models, such as Debtox, are not necessarily more complex than empirical models and they may be more biologically consistent. If this is so, then they

should be favoured if they fit the data, and empirical models used only as a default. However, in this report the empirically-based ACE two-step regression model was more successful than Debtox at predicting chronic NOECs. Further work is required to determine whether this is always the case and whether even more accurate and precise models exist. It would also be useful to compare model predictions with *true* chronic NECs, rather than the chronic NOECs that were the only data available to the report authors.

It might be argued that one of the major ways in which effluents cause mortality is by oxygen depletion due to high biochemical oxygen demand. Do the models discussed in this report adequately describe this indirect effect of chemicals in the environment? The evidence is that they do. Both early and more recent research (eg Sprague, 1969; Kooijman, 1993) suggest that the time course of lethality is similar for both chemical toxicity and oxygen starvation.

Finally, Johnson *et al.* (1996) found, as we have, that there was a close agreement between estimates of EC10 and the acute NOEC. This is not surprising, as effluent bioassays, unlike pure product toxicity tests, are constrained to produce both LC/EC and NOEC values between >0 and 100%. However, other authors have found that NOECs tend to exceed concentrations causing 10% effect in effluent bioassays (Fikslin, 1995; US EPA, 1991). Johnson *et al.* (1996) recommend that both the NOEC and EC10 are used as a Predicted No Effect Concentration (PNEC). This terminology is unfortunate. A PNEC derived in such a way is probably not an NEC, even for acute exposure periods, especially if it is simply a synonym for the EC10 (ie 10% *effect*). The PNEC is certainly *not* a chronic NEC. We believe that both the Agency and the effluent dischargers whom they regulate would benefit from the use of a summary statistic that is agreed by all to represent a safe level in the environment and has proven relevance to an agreed assessment endpoint.

5. CONCLUSIONS AND RECOMMENDATIONS

These conclusions and recommendations are derived both from this report and from the OECD report in the appendix.

- The NOEC has many disadvantages and should not be used to summarise effluent bioassay data.
- The Environment Agency should move towards a regression-based estimation procedure in which, as a minimum, the ECx at time t, model parameters, measures of error and measures of goodness of fit are reported.
- Time should be incorporated in the analytical procedures for DTA data and experimental design optimised for estimation of ECx at the last time interval, where this is both relevant to risk assessment endpoints and cost-effective.
- Procedures for the collection of data through time should be included in test protocols developed or updated in the future, where this is relevant and cost-effective.
- A modelling exercise should be undertaken to determine the optimum spacing of test concentrations, the optimum number of replicates per concentration and the optimal number of organisms per replicate for effluent bioassays.
- Predictions of chronic effects from acute effluent bioassay data sets should be tested in the field through the use of macroinvertebrate biomonitoring and *in situ* bioassay techniques (several examples of this approach may be found in Grothe *et al.*, 1996).
- The Environment Agency should initiate further study into the models that are available (both mechanistic and empirical) and which of these, if any, best meets the needs of the DTA programme. This work should extend the analyses in this report to all of the test systems recommended for DTA such as bacterial bioluminescence tests, algal growth tests and the Oyster Embryo Larval test, and to the whole range of possible endpoints, including those producing different types of discrete and continuous data.
- The relationship between different measurement endpoints such as survival, growth and reproduction should be investigated to determine whether there are predictable regularities. The use of energy budget models or life tables may be the most appropriate approach, particularly for *Daphnia* and algae. The benefit of this would be the ability to use acute data sets to predict a wider range of ecologically important chronic effects.
- A steering group should be set up by the Environment Agency to direct the mathematical, statistical and biological work required to take these recommendations forward and to provide a firm underpinning to the quantitative analysis of DTA data.

Table 3.1 Analysis of Mayer *et al.* (1992) Fish Data

Comparison	Mean y/x	SD y/x	Range y/x	Regression Coefficients		df	Intercept		Slope		r^2	r
				a	b		t	p	t	p		
Debtox and empirical chronic values												
Debtox NEC (x) & chronic NOEC (y)	0.263	0.210	0.033 - 0.704	-0.327	0.825	23	-2.221	0.0365	11.573	<0.001	0.85	0.92
Debtox NEC 95% LCL (x) & chronic NOEC (y)	1.531	5.978	0.047 - 29.523	-0.204	0.821	23	-1.93	0.0659	15.44	<0.001	0.91	0.96
Debtox 96-h EC10 (x) & chronic NOEC (y)	0.220	0.168	0.021 - 0.574	-0.228	1.018	26	-1.62	0.1173	15.49	<0.001	0.90	0.91
Debtox 96-h EC1 (x) & chronic NOEC (y)	0.227	0.174	0.023 - 0.597	-0.391	0.833	23	-2.33	0.0292	10.40	<0.001	0.82	0.91
ACE two-step linear regression and empirical chronic values												
ACE 2-step chronic EC10 (x) & chronic NOEC (y)	1.004	1.592	0.084 - 7.097	-0.061	0.0803	23	-0.40	0.6908	9.77	<0.001	0.81	0.90
ACE 2-step chronic EC1 (x) & chronic NOEC (y)	0.788	0.753	0.103 - 3.333	-0.048	0.819	23	-0.048	0.7357	10.48	<0.001	0.83	0.91
ACE 2-step chronic EC0.01 (x) & chronic NOEC (y)	1.068	1.010	0.082 - 4.610	-0.060	0.905	23	-0.605	0.5513	15.24	<0.001	0.91	0.95

ACE two-step linear regression and Debtox values												
ACE 2-step chronic EC1 (x) & Debtox NEC (y)	3.554	3.604	0.527 - 17.859	-0.337	0.985	26	4.47	<0.001	24.17	<0.001	0.96	0.98
ACE 2-step chronic EC0.01 (x) & Debtox NEC (y)	5.951	7.537	1.293 - 31.872	0.396	1.026	26	4.93	<0.001	22.16	<0.001	0.95	0.98
Toxstat 96-h ECx and empirical chronic values												
Toxstat (probit) 96-h EC50 (x) & chronic NOEC (y)	0.218	0.178	0.015 - 0.620	-0.561	0.964	14	-3.61	0.0036	14.46	<0.001	0.95	0.97
Toxstat (Spearman Karber or binomial) 96-h EC50 (x) & chronic NOEC (y)	0.231	0.173	0.015 - 0.633	-0.384	0.830	23	-2.11	0.0462	9.50	<0.001	0.80	0.89
Toxstat (probit) 96-h EC20 (x) & chronic NOEC (y)	0.273	0.218	0.017 - 0.751	-0.473	0.964	14	-3.33	0.0060	15.30	<0.001	0.95	0.98
Toxstat (probit) 96-h EC10 (x) & chronic NOEC (y)	0.308	0.240	0.019 - 0.830	0.428	0.964	14	-3.14	0.0085	15.71	<0.001	0.95	0.98
Toxstat and Debtox ECx values												
Toxstat (probit) 96-h EC50 (x) & Debtox 96-h EC50	1.146	0.234	0.907 - 1.771	0.022	1.014	14	0.62	0.5440	66.03	<0.001	0.99	0.99
Toxstat (Spearman Karber or binomial) 96-h EC50 (x) & Debtox 96-h EC50	1.135	0.213	0.893 - 1.593	0.021	1.014	26	0.82	0.4203	83.10	<0.001	0.99	0.99

Toxstat (probit) 96-h EC20 (x) & Debtox 96-h EC20	1.388	0.452	0.890 - 2.561	0.093	1.011	14	1.70	0.1117	42.17	<0.001	0.99	0.99
Toxstat (probit) 96-h EC10 (x) & Debtox 96-h EC10	1.561	0.612	0.898 - 3.144	0.134	1.010	14	2.10	0.0540	35.55	<0.001	0.99	0.99

3.2 Analysis of DTA pilot study *Daphnia* data

Comparison	Mean y/x	SD y/x	Range y/x	Regression Coefficients		df	Intercept		Slope		r^2	r
				a	b		t	p	t	p		
Debtox and empirical acute values												
Debtox NEC (x) & acute NOEC (y)	6.662	19.944	0.457 - 83.428	0.333	0.346	15	3.80	0.0017	1.76	0.0984	0.17	0.41
Debtox NEC 95% LCL (x) & acute NOEC (y)	439.64	1233.4	0.457 - 3492.1	0.434	0.171	24	11.43	<0.001	-0.686	0.4990	0.02	0.14
Debtox 96-h EC10 (x) & acute NOEC (y)	0.759	0.217	0.455 - 1.220	1.802	0.855	25	-0.038	0.970	10.01	<0.001	0.80	0.90
Debtox 96-h EC1 (x) & acute NOEC (y)	0.927	0.535	0.455 - 3.129	0.058	0.808	25	0.75	0.4642	5.31	<0.001	0.53	0.73
Toxstat 96-h ECx and empirical acute values												
Toxstat (probit) 96-h EC50 (x) & acute NOEC (y)	0.518	0.163	0.260 - 0.730	0.010	0.664	6	0.094	0.9280	4.09	0.006	0.74	0.86
Toxstat (Spearman Karber or binomial) 96-h EC50 (x) & acute NOEC (y)	0.879	0.632	0.457 - 3.181	0.117	0.610	23	1.83	0.0807	4.73	<0.001	0.49	0.70

Toxstat (probit) 96-h EC20 (x) & acute NOEC (y)	0.881	0.323	0.470 - 1.433	0.299	1.584	6	-1.59	0.1638	3.96	0.0075	0.72	0.85
Toxstat (probit) 96-h EC10 (x) & acute NOEC (y)	1.184	0.535	0.641 - 2.259	0.051	0.025	6	0.34	0.7490	2.61	0.0402	0.53	0.72
Toxstat and Debtox ECx values												
Toxstat (probit) 96-h EC50 (x) & Debtox 96-h EC50	1.015	0.246	0.720 - 1.526	0.236	0.613	6	2.02	0.0903	3.49	0.0129	0.67	0.82
Toxstat (Spearman Karber or binomial) 96-h EC50 (x) & Debtox 96-h EC50	1.599	1.490	0.695 - 6.792	0.260	0.581	24	3.17	0.0041	3.52	0.0017	0.34	0.58
Toxstat (probit) 96-h EC20 (x) & Debtox 96-h EC20	1.402	0.596	0.865 - 2.630	0.114	1.444	6	-0.39	0.7086	2.33	0.0586	0.47	0.69
Toxstat (probit) 96-h EC10 (x) & Debtox 96-h EC10	1.748	0.900	0.891 - 3.510	0.155	1.724	6	-0.39	0.7130	1.74	0.1330	0.33	0.58

REFERENCES

Abram, F.S.H. (1964) An application of harmonics to fish toxicology. *Int. J. Air Water Pollut.* **8,** 325-338.

Abram, F.S.H. (1967) The definition and measurement of fish toxicity thresholds. *Proceedings Third International Conference, Advances in Water Pollution Research*, Munich, West Germany, September 1966, pp. 75-95.

Alabaster, J.S. and Abram, F.S.H. (1965) Development and use of a direct method of evaluating toxicity to fish. *Proceedings Second International Conference, Advances in Water Pollution Research*, Tokyo, Japan, 1964, pp. 41-54.

Alderdice, D.F. and Brett, J.R. (1957) Some effects of kraft mill effluent on young Pacific salmon. *J. Fish. Res. Board Can.*, **14,** 783-795.

Bedaux, J.J.M. and Kooijman, S.A.L.M. (1994) Statistical analysis of bioassays based on hazard modeling. *J. Environ. Stat.* **1**, 303-314.

Berkson, J. (1944) Application of the logistic function to bioassay. *J. Am. Stat. Soc.* **39,** 357-365.

Bliss, C.I. (1935) The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22,** 134-167.

Bliss, C.I. and Cattell, M. (1943) Biological assay. *Ann. Rev. Physiol.* **5,** 479-539.

Bruce, R.D. and Versteeg, D.J. (1992) A statistical procedure for modeling continuous toxicity data. *Environ. Toxicol. Chem.* **11,** 1485-1494.

Buckler, D.R., Witt, A. Jr., Mayer, F.L. and Huckins, J.N. (1981) Acute and chronic effects of Kepone and Mirex on the fathead minnow. *Trans. Am. Fish. Soc.* **110,** 270-280.

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

Chen, C.W. and Selleck, R.E. (1969) A kinetic model of fish toxicity threshold. *J. WPCF* **41,** 294-308.

Chen, J.J. and Kodell, R.L. (1989) Quantitative risk assessment for teratological effects. *J. Statist. Assoc.* **84,** 966-971.

Chew, R.D. and M.A. Hamilton. 1985. Toxicity curve estimation: fitting a compartment model to median survival times. *Trans. Am. Fish. Soc.* **114,** 403-412.

Cleveland, L., Buckler, D.R., Mayer, F.L. and Branson, D.R. (1982) Toxicity of three preparations of pentachlorophenol to fathead minnows - a comparative study. *Environ. Toxicol. Chem.* **1,** 205-212.

Cox, C. (1987) Threshold dose-response models in toxicology. *Biometrics* **43,** 511-524.

Cox, D.R. (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* B **34**, 187-202.

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

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

Cripe, G.M., Goodman, L.R. and Hansen, D.J. (1984) Effect of chronic exposure to EPN and to Guthion on the critical swimming speed and brain acetylcholinesterase activity of *Cyprinodon variegatus*. *Aquatic Toxicol.* **5,** 255-266.

Crump, K.S. (1984) A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4,** 854-871.

Diamond, S.A., Newman, M.C., Mulvey, M., Dixon, P.M. and Martinson, D. (1989) Allozyme genotype and time to death of mosquitofish, *Gambusia affinis* (Baird and Girard), during acute exposure to inorganic mercury. *Environ. Toxicol. Chem.* **8,** 613-622.

Dickson, K.L., Waller, W.T., Kennedy, J.H. and Ammann, L.P. (1992) Assessing the relationship between ambient toxicity and instream biological response. *Environ. Toxicol. Chem.* **11**, 1307-1322.

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

Dwyer, W.P., Mayer, F.L., Allen, J.L. and Buckler, D.R. (1978) Chronic and simulated use-pattern exposures of brook trout (*Salvelinus fontinalis*) to 3-trifluromethyl-4-nitophenol (TFM). *Investigations in Fish Control No. 84*. US Fish and Wildlife Service, Washington, DC.

Environment Agency (1996) *The Application of Toxicity-Based Criteria for the Regulatory Control of Wastewater Discharges*. Environment Agency, Bristol, UK. 50 pp.

Environment Agency (1997) *Toxicity-Based Consents Pilot Study*. R&D Project Record P2-493/11. Environment Agency, Bristol, UK. 315 pp.

Fikslin, T.J. (1995) Development of a water quality criterion for chronic toxicity using an alternative measure of the NOEL. Poster presentation at Second SETAC World Congress, Vancouver, B.C. 5-9 November, 1995.

Finger, S.E., Little, E.F., Henry, M.G., Fairchild, J.F. and Boyle, T.P. (1985) Comparison of laboratory and field assessment of fluorene - Part 1: Effects of fluorene on the survival, growth, reproduction, and behaviour of aquatic organisms in laboratory tests. In, *Validation and Predictability of Laboratory Methods for Assessing the Fate and Effects of Contaminants in Aquatic Ecosystems*, edited by T.P. Boyle. American Society for Testing and Materials STP 865, Philadelphia, PA. pp. 120-133.

Finney, D.J. (1971) *Probit Analysis, 3rd Edition*, Cambridge University Press, Cambridge, UK.

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

Gaddum, J.H. (1953) Bioassays and mathematics. *Pharmacol. Rev.* **5,** 87-134.

Goodman, L.R., Hansen, D.J., Couch, J.A. and Forester, J. (1976) Effects of heptachlor and toxaphene on laboratory-reared embryos and fry of the sheepshead minnow. *Southeast Assoc. Game and Fish Comm.* **30,** 192-202.

Green, R.H. (1965) Estimation of tolerance over an indefinite time period. *Ecology* **46,** 887.

Grothe, D.R., Dickson, K.L. and Reed-Judkins, D.K. (1996) *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts.* SETAC Press, Pensacola, FL.

Hamelink, J.L., Buckler, D.R. , Mayer, F.L., Palawski, D.U. and Sanders, H.O. (1986) Toxicity of fluoridone to aquatic invertebrates and fish. *Environ. Toxicol. Chem.* **5,** 87-94.

Hansen, D.J. and Cripe, G.M. (1984) *Interlaboratory Comparison of the Early Life-Stage Toxicity Test Using the Sheepshead Minnow* (Cyprinodon variegatus). EPA-600/X-84-081. US Environmental Protection Agency, Gulf Breeze, FL.

Hansen, D.J. and Parrish, P.R. (1977) Suitability of sheepshead minnows (*Cyprinodon variegatus*) for lifecycle toxicity tests. In, *Aquatic Toxicology and Hazard Evaluation*, edited by F.L. Mayer and J.L. Hamelink. American Society for Testing and Materials STP 634, Philadelphia, PA. pp 117-126.

Hansen, D.J., Schimmel, S.C. and Forester, J. (1977) Endrin: effects on the entire life cycle of a saltwater fish, *Cyprinodon variegatus*. *J. Toxicol. Environ. Hlth.* **3,** 721-733.

Hartley, H.O. and Sielken, R.L. Jr. (1977) Estimation of 'safe doses' in carcinogenic experiments. *Biometrics* **33,** 1-30.

Heming, T.A., A. Sharma and Y. Kumar. (1989). Time-toxicity relationships in fish exposed to the organochlorine pesticide methoxychlor. *Environ. Toxicol. Chem.* **8,** 923-932.

Herbert, D.W.M. and Shurben, D.S. (1964) The toxicity of fluoride to rainbow trout. *Water and Waste Treatment* Sept/Oct 1994, pp. 141-142.

Hey, E.N. and Hey, M.H. (1960) The statistical estimation of a rectangular hyperbola. *Biometrics* **16**, 606-617.

Hoekstra, J.A. (1991) Estimation of the LC50, a review. *Environmetrics* **2**, 139-152.

Hoekstra, J.A. and van Ewijk, P.H. (1993) Alternatives for the no-observed effect level.

Environ. Toxicol. Chem. **12,** 187-194.

Johnson, I., Wharfe, J., Tinsley, D., Boumphrey, R. and Forrow, D. (1996) *Toxicity-Based Consents Pilot Study*. R&D Technical Report P23, Environment Agency, Bristol, UK. 35 pp.

Kalbfleisch, J.D., Krewski, D.R. and van Ryzin, J. (1983) Dose-response models for time-to-response toxicity data. *Can. J. Statist.* **11,** 25-49.

Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457-481.

Kerr, D.R. and Meador, J.P. (1996) Modeling dose response using generalized linear models. *Environ. Toxicol. Chem.* **15,** 395-401.

Kenaga, E.E. (1982) Predictability of chronic toxicity from acute toxicity of chemicals in fish and aquatic invertebrates. *Environ. Toxicol. Chem.* **1,** 347-358.

Kooijman, S.A.L.M. (1981) Parametric analyses of mortality rates in bioassays. *Wat. Res.* **15**, 107-119.

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

Kooijman, S.A.L.M. (1993) *Dynamic Energy Budgets in Biological Systems: Theory and Applications in Ecotoxicology*. Cambridge University Press.

Kooijman, S.A.L.M. (1996) An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* **75,** 310-316.

Kooijman, S.A.L.M. and Bedaux, J.J.M. (1996a) Analysis of toxicity tests on fish growth. *Wat. Res.* **30**, 1633-1644.

Kooijman, S.A.L.M. and Bedaux, J.J.M. (1996b) Analysis of toxicity tests on *Daphnia* survival and reproduction. *Wat. Res.* **30**, 1711-1723.

Kooijman, S.A.L.M. and Bedaux, J.J.M. (1996c) Some statistical properties of estimates of no-effect concentrations. *Wat. Res.* **30**, 1724-1728.

Kooijman, S.A.L.M. and Bedaux, J.J.M. (1996d) *The Analysis of Aquatic Toxicity Data*. VU University Press, Amsterdam. 149 pp.

Kooijman, S.A.L.M., Hanstveit, A.O. and Nyholm, N. (1996) No-effect concentrations in algal growth inhibition tests. *Wat. Res.* **30**, 1625-1632.

Laskowski, R. (1995) Some good reasons to ban the use of NOEC, LOEC and related concepts in ecotoxicology. *Oikos* **73,** 140-144.

Lee, G., Ellersieck, M.R., Mayer, F.L. and Krause, G.F. (1995) Predicting chronic lethality of chemicals to fishes from acute toxicity test data: multifactor probit analysis. *Environ.*

Toxicol. Chem. **14**, 345-349.

Lloyd, R. (1960) The toxicity of zinc sulphate to rainbow trout. *Ann. appl. Biol.* **48**, 84-94.

Marcus, M.D. amd McDonald, L.L. (1992) Evaluating the statistical bases for relating receiving water impacts to effluent and ambient toxicities. *Environ. Toxicol. Chem.* **11**, 1389-1402.

Matida, Y. (1960) Study on the toxicity of agricultural control chemicals in relation to fisheries management. No. 3. A method for estimating threshold value and a kinetic analysis of the toxicity curve. *Bull. Freshwat. Fish. Res. Lab. Tokyo* **9**, 1-12.

Mayer, F.L., Adams, W.J., Finley, M.T., Michael, P.R., Mehrle, P.M. and Saeger, V.W. (1981) Phosphate ester hydraulic fluids: an aquatic environmental assessment of Pydrauls 50E and 115E. In, *Aquatic Toxicology and Hazard Assessment*, edited by D.R. Branson and KL. Dickson. American Society for Testing and Materials STP 737, Philadelphia, PA. pp 103-123.

Mayer, F.L., Mehrle, P.M. and Dwyer, W.P. (1975a) Toxaphene effects on reproduction, growth and mortality of brook trout. EPA-600/3-75-013. United States Environmental Protection Agency, Duluth, MN.

Mayer, F.L., Mehrle, P.M. and Dwyer, W.P. (1975b) Toxaphene: chronic toxicity to fathead minnows and channel catfish. EPA-600/3-77-069. United States Environmental Protection Agency, Duluth, MN.

Mayer, F.L., Mayer, K.S. and Ellersieck, M.R. (1986) Relation of survival to other endpoints in chronic toxicity tests with fish. *Environ. Toxicol. Chem.* **5**, 737-748.

Mayer, F.L., Krause, G.F., Ellersieck, M.R. and Lee, G. (1991) Statistical approach to predicting chronic toxicity of chemicals to fishes from acute toxicity test data. PB92-169655. National Technical Information Service, Springfield, VA.

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 analysis. *Environ. Toxicol. Chem.* **13**, 671-678.

Mayer, F.L., Sun, K., Lee, G., Ellersieck, M.R. and Krause, G.F. (1996) *User Guide: Acute to Chronic Estimation (ACE)*. United States Environmental Protection Agency, Gulf Breeze, FL.

Mount, D.I. and Stephan, C.E. (1967) A method for establishing acceptable limits for fish - malathion and the butoxyethanol ester of 2,4-D. *Trans. Am. Fish. Soc.* **96**, 185-193.

Nelson, W. (1982) *Applied Life Data Analysis*, Wiley, New York.

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

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

Newman, M.C., Diamond, S.A., Mulvey, M. and Dixon, P. (1989) Allozyme genotype and time to death of mosquitofish, *Gambusia affinis* (Baird and Girard) during acute toxicant exposure: a comparison of arsenate and inorganic mercury. *Aquat. Toxicol.* **15,** 141-156.

Newman, M.C., Keklak, M.M. and Doggett, M.S. (1994) Quantifying animal size effects on toxicity: a general approach. *Aquat. Toxicol.* **28,** 1-12.

Noppert, F., Leopold, A. and van der Hoeven, N. (1994) *How to Measure No Effect: Towards a New Measure of Chronic Toxicity in Ecotoxicology.* BKH Consulting Engineers, Delft, The Netherlands. 44 pp.

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

Oris, J.T. and Bailer, A.J. (1997) Equivalence of concentration-response relationships in aquatic toxicology studies: testing and implications for potency estimation. *Environ. Toxicology Chem* **16**, 2204-2209.

Pack, S. (1993) *A Review of Statistical Data Analysis and Experimental Design in OECD Aquatic Toxicology Test Guidelines.* Report to the Organisation for Economic Cooperation and Development, Paris. 42 pp.

Parrish, P.R., Schimmel, S.C., Hansen, D.J., Patrick, J.M. and Forester, J. (1976) Chlordane: effects on several estuarine organisms. *J. Toxicol. Environ. Hlth.* **1,** 485-494.

Prentice, R.L. (1976) A generalization of the probit and logit methods for dose response curves. *Biometrics* **32,** 761-768.

Powers, E.B. (1917) The goldfish (*Carassius carassius*) as a test animal in the study of toxicity. *Illinois Biol. Monogr.* **4,** 127-193.

Rand, G.M., Wells, P.G. and McCarty, L.S. (1995) Introduction to aquatic toxicology. In, *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment, 2nd Edition*, edited by G.M. Rand. Taylor and Francis, Washington, D.C.

Shirazi, M.A. and Lowrie, L. (1988) Comparative toxicity based on similar asymptotic endpoints. *Arch. Environ. Contam. Toxicol.* 17, 273-280.

Skalski, J.R. (1981) Statistical inconsistencies in the use of no-observed-effect-levels in toxicity testing. In, *Aquatic Toxicology and Hazard Assessment: Fourth Conference*, edited by Branson, D.R. and Dickson, K.L., American Society for Testing and Materials, Philadelphia, PA. pp. 377-387.

Sprague, J.B. (1969) Measurement of pollutant toxicity to fish. I. Bioassay methods for acute toxicity. *Water Res.* **3,** 793-821.

Stephan, J.B. (1977) Methods for calculating an LC50. In, *Aquatic Toxicology and Hazard Evaluation*, edited by Mayer, F.L. and Hamelink, J.L., American Society for Testing and Materials, Philadelphia, PA. pp. 65-84.

Stephan, C.E. and Rogers, J.W. (1985) Advantages of using regression to calculate results of chronic toxicity tests. In, *Aquatic Toxicology and Hazard Assessment: Eighth Symposium*, edited by Bahner, R.C. and Hansen, D.J., American Society for Testing and Materials, Philadelphia, PA. pp. 328-338.

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. *Environ. Toxicol. Chem.* **14,** 1745-1752.

Suter, G.W. II (1993) *Ecological Risk Assessment*, Lewis Publishers, Chelsea, MI.

Suter, G.W. II, Rosen, A.E., Linder, E. and Parkhurst, D.F. (1987) Endpoints for responses of fish to chronic toxic exposures. *Environ. Toxicol. Chem.* **6,** 793-809.

West, Inc. (1994) *Toxstat 3.4,* Western Ecosystems Technology Inc, WY.

Unistat Ltd. (1994) *Unistat Statistical Package Version 3.0 for MS Windows: User's Guide,* Unistat Ltd, London.

US EPA (1981) *Acephate, aldicarb, carbophenothion, DEF, EPN, ethoprop, methyl parathion and phorate: their acute and chronic toxicity, bioconcentration potential and persistence as related to marine environments*. EPA-600/4-81-023. United States Environmental Protection Agency, Gulf Breeze, FL.

US EPA (1991) *Technical Support Document for Water Quality Based Toxics Control*. United States Environmental Protection Agency Office of water, Washington, D.C., EPA 505/2-90-001.

Weibull, W. (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics* **18**, 293-297.

Whitehouse, P., Crane, M., Redshaw, C.J. and Turner, C. (1996) Aquatic toxicity tests for the control of effluent discharges in the UK - the influence of test precision. *Ecotoxicology* **5,** 155-168.

Widianarko, B. and van Straalen, N. (1996) Toxicokinetics-based survival analysis in bioassays using nonpersistent chemicals. *Environ. Toxicol. Chem.* **15,** 402-406.

Zar, J.H. (1984) *Biostatistical Analysis*, Second Edition. Prentice Hall, NJ.

APPENDIX OECD BRAUNSCHWEIG DRAFT REPORT

APPENDIX

DRAFT REPORT OF THE

OECD WORKSHOP ON STATISTICAL ANALYSIS OF AQUATIC TOXICITY DATA, BRAUNSCHWEIG, GERMANY, 15-17 OCTOBER 1996

December 1996

Executive Summary

Following a decision taken by the National Co-ordinators of the Test Guideline Programme and the HAAB in their joint session in December 1995, an OECD Workshop on Statistical Analysis of Aquatic Ecotoxicity Data was held in Germany on 15-17 October 1996. The meeting was hosted by the BBA in Braunschweig and was chaired by Dr. Arno Lange from the German UBA.

The objectives of the workshop were to:

(i) review the options available for the analysis of data from ecotoxicity tests;
(ii) compare their advantages and disadvantages;
(iii) recommend (a) the most appropriate approach for deriving a summary parameter(s) which has scientific validity, and (b) further work for OECD and/or others, as appropriate.

In a series of breakout and plenary sessions, the workshop discussed statistical data analysis appropriate for single-species chronic/subchronic studies using a number of test concentrations. Aquatic tests served as a basis for these discussions although the issues addressed may be similar for ecotoxicity tests in general. Background documents had been prepared on the following main existing approaches of data analysis for such tests:

- Analysis of Variance/Hypothesis Testing ("ANOVA/NOEC approach")
- Regression analysis (based on empirical models)
- Mechanistic modelling (theory-based)

The workshop concluded that the NOEC as the main summary parameter of aquatic ecotoxicity tests is scientifically inappropriate for a number of reasons and should therefore be phased out. It was recommended that OECD should move towards a regression-based estimation procedure. The time course of effects should be incorporated in the analytical procedures. OECD should initiate a study into the dynamic regression models that are available (both mechanistic and empirical) in order to select those which best meet OECD's needs. The study should also address the issue of appropriate values of 'x' for ECx and the optimal experimental designs. A steering group should be set up to direct the mathematical, statistical and biological work required to take the workshop recommendations forward. This group should include representatives from the appropriate scientific and regulatory communities.

Introduction

Background

Within OECD countries, a number of aquatic ecotoxicity test guidelines are used to assess the potential effects of chemicals (including pesticides) on aquatic organisms. For the use in hazard/risk assessment schemes, summary parameters (e.g., the LC50 and the NOEC) are established by statistical methods. This statistical evaluation plays a major role in developing the test guidelines since the experimental design is crucial for the statistical method that can be applied, and both together are central for developing tests that produce high-quality data with a minimum use of resources and test organisms.

In 1992, the National Coordinators for the OECD Test Guidelines Programme decided to have a review of existing and draft aquatic toxicity test guidelines with respect to test design and statistical data analysis. The "Review of Statistical Data Analysis and Experimental Design in OECD Aquatic Toxicology Test Guidelines" (Annex 10) was prepared by Dr. Simon Pack in 1993 and widely circulated and discussed. The review and the recommendations made were broadly well-received and appreciated. Other activities of the scientific community on this issue included workshops in The Hague (1994) and in London (1996).[1]

All these activities concluded that the NOEC is inappropriate as a summary measure of toxicity. Replacing the NOEC as suggested has implications for the test designs as well as for hazard/risk assessment. Hence, within the context of OECD work, both the OECD Test Guidelines and the Hazard/Risk Assessment Programmes became involved, and the relevant bodies (the National Co-ordinators of the Test Guideline Programme and the HAAB) decided to have a joint workshop to proceed the issue further.

The OECD Workshop on Statistical Analysis of Aquatic Ecotoxicity Data was held in Germany on 15-17 October 1996. The meeting was hosted by the BBA in Braunschweig and was chaired by Dr. Arno Lange from the German UBA. There were 52 participants from 15 Member countries and industry (Annex 1).

Objectives

The objectives of the workshop were to:

(i) review the options available for the analysis of data from ecotoxicity tests;
(ii) compare their advantages and disadvantages;
(iii) recommend (a) the most appropriate approach for deriving a summary parameter(s) which has scientific validity, and (b) further work for OECD and/or others, as appropriate.

---

[1]Noppert, F., N. Van der Hoeven and A. Leopold.. 1994. How to measure no effect : towards a new measure of chronic toxicity in ecotoxicology. Workshop Report of the Netherlands Working Group on Ecotoxicology. The Hague
Chapman, P.F., M. Crane, J.A. Wiles, F. Noppert and E.C. McIndoe (eds.). 1996. Asking the Right Questions: Ecotoxicology and Statistics. Report of a Workshop Held at Royal Holloway University of London, Surrey, U.K. SETAC-Europe

Focus

The workshop focused on approaches to data analysis appropriate for single-species chronic/subchronic *aquatic tests* using a number of test concentrations, although it was recognized that the discussion might also be relevant to the analysis of data from ecotoxicity tests in general. The implications of the different statistical approaches on test design were also considered.

Workshop structure and discussion topics

The workshop was organised around a series of plenary sessions and three working groups (Annex 2). These working groups all addressed the same issues in parallel sessions and reported on progress during plenary sessions. Discussion topics were framed as questions and are listed below.

Session 1:

- Why are OECD ecotoxicity tests performed ?
- What kind of information do we want from the tests ?
- Are we happy with the current statistical practices ?

Session 2:

- Review and comparison of the different approaches to data analysis:
  - ◊ Should the NOEC be retained ?
  - ◊ Which other analytical technique could replace the NOEC ?
  - ◊ What type of information (statistical summary parameter; test endpoint) do we want from the new approach ?
- What work needs to be done with respect to selection of statistical approach ?

The individual working group reports are given in Annexes 3, 4 and 5.

Background documents

Documents describing the three main approaches to data analysis, including an assessment of their strengths and weakness, were prepared and distributed in advance of the workshop:

- A Discussion of the NOEC/ANOVA Approach to Data Analysis (Annex 6)
- Alternatives to the NOEC Based on Regression Analysis (Annex 7)
- Dynamic measures for ecotoxicity (Annex 8)
- The Dynamic Energy Budget (DEB) model (Annex 9)

Summary of Plenary Discussions

Plenary Rapporteurs: Peter Chapman (UK), and Mark Crane (UK)

This section summarises the main outcomes of the workshop. The discussion issues, together with a summary of the views put forward, and a list of recommendations arising out of the workshop, are listed below. There was a high degree of agreement on all of the issues, and, where there was disagreement or lack of consensus, this is highlighted. A more detailed account of the proceedings of each working group can be found in their individual reports (see Annexes 3 - 5).

1. Why are OECD Ecotoxicity Tests Performed?

Ecotoxicity tests are performed to evaluate the toxicity of chemicals in order to predict their potential effects on natural populations. These tests provide information which is used in the registration / notification of new chemicals and the assessment of older chemicals, including pesticides and biocides. The results from these tests are used in the classification and labelling of chemicals and contribute the 'effects' component to a risk assessment. They may also be used to predict adverse effects in the event of an accident.

OECD Test guidelines are primarily developed for the above reasons, but may also be used in other situations, including the bioassay of environmental samples and fundamental research.

2. What Kind of Information Do We Need from OECD Ecotoxicity Tests?

The workshop agreed that toxicity tests should provide information that is accurate and precise, and that permits easy interpretation by the non-expert. Ideally, information on the time course of effects should be integrated with information on the concentration-response curve. The information should be of 'biological relevance', although there was no consensus on whether this should refer simply to the types of measurements taken (e.g., survival, growth and reproduction are usually considered as relevant parameters because changes in them can affect population abundance), or to 'ecological relevance'.

It was recognised that the specific type of information required from a test will largely depend on the way in which it is extrapolated to the natural world. There was no consensus on whether 'classification' and 'risk assessment' demanded different information and analyses, or whether classification is simply a point on the road to risk assessment and which uses similar data and analyses. However, classification will normally use only acute lethal data, while risk assessment will often use both lethal and sublethal data.

3. Are We Happy with Current Statistical Practices?

There was virtually unanimous agreement that current practices were unsatisfactory and there was a great deal of consistency in the views put forward. The list below is a comprehensive summary of all views raised.

- There is a concern that the NOEC may not be sufficiently protective because of the danger of false negatives.
- The statistical methods are suboptimal.

- OECD Test guidelines contain insufficient information on statistical techniques.
- Data are wasted in the determination of values such as the NOEC.
- NOECs are leading to misunderstandings and misinterpretations.
- Current summary statistics cannot be linked to population models.
- There are no statements of biological significance, only statistical significance (there was no consensus on this issue because 'biological significance' appears to mean different things to different people).
- More effective use should be made of test animals.
- Results are often imprecise.
- Biologically relevant covariates are not considered.

4. What Should OECD Tests Look Like in Future?

It was agreed that new testing frameworks should not be developed that (a) exclude results from tests performed within the current framework, or (b) remove all flexibility in approach. However, there was common agreement that certain improvements are desirable:

- The danger of false negatives should be reduced.
- Tests should focus on biologically significant endpoints (although a definition needs to be agreed first).
- We should be able to link test results to predictive ecological/biological models.
- Biologically important covariates should be included, where this is appropriate.
- Better use should be made of all test measurements.
- Test guidelines should make explicit recommendations on appropriate statistical analytical techniques.
- Summary statistics and parameters should be capable of being interpreted by non-statisticians and decision makers.

5. Comparison of the Different Approaches to Data Analysis

5.1 Hypothesis testing to determine an NOEC

Advantages

- The NOEC is easy to understand (but regression is also easy to understand). However, only one model is available versus many potential models to choose if OECD moves to regression.
- It can always be determined when regression models fail (but its determination under these circumstances may not always be 'reasonable').
- Occasionally there are large confidence intervals for regression analyses (but there are none at all for NOECs).

Disadvantages

- The NOEC itself is statistically unfounded. (The review report [Annex 10] by Pack gives a detailed explanation of this point.)
- Hypothesis testing in general is not well suited to the type of data obtained from most toxicity tests (with the possible exception of limit tests).

5.2. Static regression

Static regression models, in which a model is fitted to measurements taken at a single fixed time, and which therefore does not include a time component, do not use all of the data in an efficient manner. However such models may be politically acceptable 'stepping stones' to dynamic analyses. Static regression models are the only option for those endpoints which are assessed only once.

5.3. Dynamic regression

Time of exposure should be incorporated into the analysis of data, where possible. However, many different dynamic regression models are available, each with their own assumptions. Criteria are needed for the appropriate selection from the available mechanistic and empirical models. Best reasonable fit should be the most important criterion for model selection. Also the model, whilst being as complex as necessary, should be as simple as possible. Dynamic models are not necessarily more complex than static models (there can be less parameters). It may be that empirical models are less biologically consistent than mechanistic models (eg increases in survival may occur over time). Mechanistic models should be favoured if they fit the data; empirical models should be used as a fallback.

In the short to medium term there is no need to modify test protocols in major ways. However, minor modifications may produce a better result. It may also be useful to collect data at intermediate time intervals in some tests, such as the fish growth test, although there are more cost implications in this proposal.

6. Concerns About Moving Away From the NOEC

Concerns were expressed about deciding to abandon the NOEC before alternative methods have been identified and evaluated for their implications on ecotoxicity test designs. It is possible that specific methods will be needed for each species and ecotoxicological endpoint. Where several methods exist, guidance for their selection for regulatory use would be required.

Workshop Conclusions and Recommendations

1. The NOEC should be phased out as a summary of toxicity.

2. OECD should move towards a regression-based estimation procedure in which, as a bare minimum, the following should be reported: model parameters plus measures of error and goodness of fit; ECx,t; important biological parameters; parameters describing the time course of effects.

3. Time should be incorporated in the analytical procedures for OECD toxicity test data and experimental designs should be optimised for estimation of ECx at the last time interval (where this is both relevant and cost-effective).

4. If different models are equivalent and give adequate fits to data, and if assumptions are valid, mechanistic models are preferred over empirical models.

5. Procedures for the collection of data through time should be included in test protocols that are developed or updated in the future (where this is relevant and cost-effective).

6. OECD should initiate a study into the dynamic regression models that are available (both mechanistic and empirical) and which of these, if any, best meets OECD's needs. This study should include discussion of the following:

   — what statistical summaries are robust and should be reported to regulatory authorities?

   — other biological estimates or parameters that should be reported.

   — a sensitivity analysis of test design to justify an appropriate ECx (data should be reanalysed to determine the precision associated with low ECx values and the optimal experimental design).

   — whether a different x is required for different tests because of the different levels of precision achievable in each.

   — an analysis of the advantages and disadvantages to risk assessment of a move to dynamic regression approaches for different organisms and endpoints.

7. A steering group should be set up to direct the mathematical, statistical and biological work required to take the workshop recommendations forward. This group should include representatives from the appropriate scientific and regulatory communities. The group could gather data from the testing community via a questionnaire.

8. The report of this workshop, including background documents and the 1993 review report by Pack, should be published as an OECD Monograph.

ANNEX 1

OECD WORKSHOP ON STATISTICAL ANALYSIS OF AQUATIC ECOTOXICITY DATA

Braunschweig, Germany, 15th-17th October 1996

AUSTRIA
AUTRICHE

Norbert Bomatowicz
Österreichisches Forschungszentrum
Seibersdorf
A-244 Seibersdorf
AUTRICHE

Tel: 43-1-2254-780-3540
Fax: 43-1-2254-780-3653
E-mail: bomatowicz@zdfzs.arcs.co.at

Britta Grillitsch
Veterinärmedizinische Universität Wien,
Josef Baumanngass 1,
A-1201 Wien
AUTRICHE

Tel: 43-1-250-77-4604/1
Fax: 43-1-250-72-4709
E-mail:britta.grillitsch@vu-wien.ac.at

BELGIUM
BELGIQUE

Katrien Delbeke
LISEC
Craenevenne 140
B-3600 GENK
BELGIQUE

Tel: 32-89-36-2791
Fax: 32-89-35-5805
E-mail:

Colin Janssen
Lab for -Biological Research in Aquatic Pollution
University of Ghent
Plateaustraat 22
B-9000 GENT
BELGIQUE

Tel: 32-9-264-3775
Fax: 32-9-264-4199
E-mail: colin.jansen@rug.ac.be

Isabelle Halleux
ISSEP
Rue du Chera 200
B-4000 Liège
BELGIQUE

Tel: 32-4-2527150
Fax: 32-4-2424665
E-mail:

CANADA
CANADA

Glen Atkinson
attn. Mark Lewis
Commercial Chemicals Evaluation Branch
Environment Canada
Place Vincent Massey, 14th floor
351 St. Joseph Blvd.
Hull, Quebec K1A OH3
CANADA

Tel: 1-613-232-4621
Fax: 1-819-953-4936
E-mail: g-f.atkinson@sympatico.ca

Peter Delorme
PMRA
Tupper Building
2250 Riverside Drive
Ottawa, Ontario K1A 0K9
CANADA

Tel: 1-613 736-3729
Fax: 1-613 736-3710
E-mail: pdelorme@pmra.hwc.ca

Dwayne Moore
The Cadmus Group
411 Roosevelt Avenue, Suite 204
Ottawa, Ontario K2A 3X9
CANADA

Tel: 1-613-761-1464/1568
Fax: 1-613-761-7653
E-mail: moored@ibm.net

DENMARK
DANMARK

Claus Hansen
Danish EPA, Pesticides Division
Strandgade 29,
DK-1401 Copenhagen
DANEMARK

Tel: 45-32-66-0100
Fax: 45-32-66-0479
E-mail:

Helle Holst
DTU, Dept. for Mathematical Modelling
Building 321
DK-2800 Lyngby
DANEMARK

Tel: 45-45-25-3357
Fax: 45-32-66-0479
E-mail: hh@imm.dtu.dk

Gerard Jagers op Akkerhuis
National Environmental Research Institute
P.O Box 314
Vejlsovej 25
DK- 8600 Silkeborg
DANEMARK

Tel: 45-89-20-15-72
Fax: 45-89-20-14-14
E-mail: gja@gmu.dk

Niels Nyholm
DTU, Institute of Environmental Science
and Engineering
Building 113
DK-2800 Lyngby
DANEMARK

Tel: 45-45-25-1471
Fax: 45-45-93-2850
E-mail: nn@imt.dvt.dk

FINLAND
FINLANDE

Hannu Braunschweiler
Finnish Environment Institute
P.O. Box 140
00251 Helsinki
FINLANDE

Tel: 358-9-4030-0538
Fax: 358-9-4030-0591
E-mail: hannu.braunschweiler@vyh.fi

FRANCE
FRANCE

Jean-François Férard
Centre des Sciences de l'Environnement
1, rue des Récollets
57000 Metz
FRANCE

Tel: 33-3-8775-8180/81
Fax: 33-3-8775-8189
E-mail:

Eric Vindimian
INERIS
Parc technologique Alata
B.P 2
60550 Verneuil en Halatte
FRANCE

Tel: 33-4455-6827
Fax: 33-4455-6655
E-mail:

GERMANY
ALLEMAGNE

Mr Arno W. Lange (Chairman)
Umweltbundesamt
Bismarckplatz 1
Postfach 330022
D-14191 Berlin
ALLEMAGNE

Tel: 49-30-8903 3110
Fax: 49-30-8903 3903
E-mail: arno.lange@uba.de

Toni Ratte
Rheinisch-Westfälische
Technische Hochschule Aachen
Lehrstuhl für Biologie V
Worringerweg 1
52056 Aachen
ALLEMAGNE

Tel: 49-241-806-680
Fax: 49-241-888-8182
E-mail: ratte@rwth-aachen.de

Sabine Martin
Umweltbundesamt
IV 2.4
Postfach 330022
14193 Berlin
ALLEMAGNE

Tel:
Fax:
E-mail:

Martin Streloke
BBA
Messeweg 11/12
D-38104 Braunschweig
ALLEMAGNE

Tel: 49-531-299-3609
Fax: 49-531-299-3005
E-mail:

ITALY
ITALIE

Silvia Marchini
Laboratorio di Tossicologia
Comparata e Ecotossicologia
Istituto Superiore di Sanità
Viale Regina Elena, 299
00161 Roma
ITALIE

Tel: 39-6-4990-2786
Fax: 39-6-4440-140
E-mail: lansi@iss.it

THE NETHERLANDS
PAYS-BAS

Jacques Bedaux
Department of Theoretical Biology
Vrije Universiteit
De Boelelaan 1087
1081 HV Amsterdam
PAYS BAS

Tel: 31-20-444-7128
Fax: 31-20-444-7123
E-mail: bedaux@bio.vu.nl

Rinus Bogers
Notox bv.
P.O. Box 3476
5203 DL s'Hertogenbosch
PAYS BAS

Tel: 31-73-641-9575
Fax: 31-73-641-8543
E-mail:

Cees J. van Leeuwen
Quality Division / Risk Assessment and Environmental
Ministerie van VROM
Postbus 30945
NL-2500 Gravenhage
PAYS-BAS

Tel: 31-70-339-4943
Fax: 31-70-339-1314
E-mail:

NORWAY
NORVEGE

Erlend Spikkerud
Landbrukstilsynet
Norwegian Agricultural Inspection Service
P.O. Box 3
N-1430 ÅS
NORVEGE

Tel: 47-64-94-4400
Fax: 47-64-94-4410
E-mail:

SPAIN
ESPAGNE

Enrique Andreu Moliner
Departamento de Biología Animal
Laboratory of Ecotoxicology
Universidad de Valencia
Dr. Moliner, 50
E-46100-Bujasot (Valencia)
ESPAGNE

Tel: 34-6-386-4676
Fax: 34-6-386-4372
E-Mail: enrique.andreu@uv.es

SWEDEN
SUEDE

Björn Dahl
Astra-Hässle AB
S-431 83 Mölndal
SUEDE

Tel: 4631-776-2950
Fax: 4631-776-3787
Email:bjorn.dahl@hassle.sc.astra.com

Lars Lindqvist
National Chemicals Inspectorate
P.O. Box 1384
S-171 27 Solna
SUEDE

Tel: 468-730 6836
Fax: 468-735-7698
E-mail: larsl@kemi.se

SWITZERLAND
SUISSE

Roland D. Fisch
Ciba-Geigy AG
Mathematical Applications/IS 2.4
R-1008.Z230
CH-4002 Basel
SUISSE

Tel: 41-61-697-6452
Fax: 41-61-697-8973
E-mail: wrfi@chbs.ciba.com

UNITED KINGDOM
ROYAUME UNI

Mark Crane
Division of Biology
School of Biological Sciences
Royal Holloway University of London
Egham, Surrey TW20 0EX
ROYAUME-UNI

Tel: 44-1784-443372
Fax: 44-1784-470756
E-mail: m.crane@rhbnc.ac.uk

John S. Fenlon
Horticulture Research International
Wellesbourne, Warwick CV35 9EF
ROYAUME-UNI

Tel: 44-1789-470382
Fax: 44-1789-470552
E-mail: john.fenlon@hri.ac.uk

Andrew Riddle
Zeneca Ltd.,
Brixham Environmental Laboratory
Freshwater Quarry,
Brixham, Devon TQ5 8BA
ROYAUME-UNI

Tel: 44-1803-882882
Fax: 44-1803-882974
E-mail:riddle@bx1vax.zeneca.com

Tim Sparks
Institute of Terrestrial Ecology
Monks Wood, Abbots Ripton
Huntingdon, Cambridgeshire PE17 2LS
ROYAUME-UNI

Tel: 44-1487-773381
Fax: 44-1487-773467
E-mail: ths@wpo.nerc.ac.uk

UNITED STATES
ETATS-UNIS

Richard G. Clements
US-EPA
Office of Pollution Prevention and Toxics (7403)
401 m Street S.W.
Washington DC 20460
ETATS-UNIS

Tel: 1-202-260-5270
Fax: 1-202-260-1283
E-mail:clements.dick@epamail.epa.gov

David Farrar
US-EPA
Office of Pollution Prevention and Toxics (7403)
401 m Street S.W.
Washington DC 20460
ETATS-UNIS

Tel: 1-703-305-5721
Fax:
E-mail: farrar.david@epamail.epa.gov

Michael C. Newman
University of Georgia
P.O Drawer E
Aiken, SC 29802
ETATS-UNIS

Tel: 1-803-725-5746
Fax: 1-803-725-3309
E-mail: newman@srel.edu

BIAC

Michael C. Harrass
Amoco Corporation
Mail Code PO619U9
130 E. Randolph Drive
Chicago, IL 60601
ETATS-UNIS

Tel: 1-312-856-5116
Fax: 1-312-856-7584
E-mail:mjharrass@amoco.com

Roland Maisch
BASF AG
DUU/OO - Z 570
D-67056 Ludwigshafen
ALLEMAGNE

Tel: 49-621 60 58041
Fax: 49-691 60 58043
E-mail:roland.maisch@duu.x400.basf-ag.de

Joanna Jaworska
Environmental Toxicologist
N.V. Procter & Gamble ETC
Temselaan 100
B-1853 Strombeek-Bever
BELGIQUE

Tel: 32-2-456-2076
Fax: 32-2-456-2845
E-mail: jaworska.j@pg.com

ECETOC

Roger Van Egmond
Unilever Research / Port Sunlight Laboratory
Quarry Road East
Bebington,
Wirral Merseyside L63 3JW
ROYAUME-UNI

Tel: 44-151-471-3917
Fax: 44-151-471-1847

Kathleen M. Stewart
ZENECA
Brixham Environmental Laboratory
Freshwater Quarry
Brixham, Devon TQ5 8BA
ROYAUME-UNI

Tel: 44-1803-882882
Fax: 44-1803-282974
E-mail:

Lisa Tattersfield
Shell Research Ltd,
Thornton Research Centre
P.O. Box 1
Chester CH1 3S11
ROYAUME-UNI

Tel: 44-151-373-5019
Fax: 44-151-373-5845
E-mail: l.tattersfield@msmail.trctho.simis.com

GIFAP/ECPA

Kees Romijn
Rhône-Poulenc Agro
355 rue Dostoievski
BP 153
F-06903 Sophia Antipolis
FRANCE

Tel: 33-4 9294-3453
Fax: 33-4 9365-3924
E-mail:

STEERING COMMITTEE
COMITE DIRECTEUR

Pascal Isnard
Rhone-Poulenc Industrialisation
C.R.I.T.
24, avenue Jean Jaurès
69153 Decines Charpieu Cedex
FRANCE

Tel: 33-7205-2568
Fax: 33-7205-2350
E-mail:

Gerhard Joermann
BBA
Messeweg 11/12
D-38104 Braunschweig
ALLEMAGNE

Tel: 49-531-299-3613
Fax: 49-531-299-3005
E-mail: g.joermann@bba.de

Bas Kooijman
Vrije Universiteit
de Boelelaan 1087
1081 HV Amsterdam
PAYS BAS

Tel: 31-20-444-7130
Fax: 31-20-444-7123
E-mail: bas@bio.vu.nl

Reinhard Meister
TFH Berlin
Fachbereich 2
Luxemburger Str. 10
13353 Berlin
ALLEMAGNE

Tel: 49-30-450-2213
Fax: 49-30-450-4201
E-mail: meister@tfh-berlin.de

José Tarazona
CISA-INIA
28130 Valdeolmos
Madrid
ESPAGNE

Tel: 34-1-620-2300
Fax: 34-1-620-2247
E-mail:

Les Touart
U.S. EPA
OPP (7507C)
401 M Street, SW
Washington, DC 20460
ETATS-UNIS

Tel: 1-703-305-6134
Fax: 1-703-305-6309
E-mail: Touart.les@epamail.epa.gov

Peter Chapman
Zeneca Agrochemicals
Jeallott's Hill Research Station
Bracknell
Berkshire RG12 6EY
ROYAUME-UNI

Tel: 44-134-441-4694
Fax: 44-134-441-4853
E-mail:Peter.p.f.chapman@gbjha.zeneca.com

Simon Pack
Procter & Gamble Pharmaceuticals
Lovett House, Lovett Road
Staines
Middlesex TW18 3AZ
ROYAUME-UNI

Tel: 44-1784-49-5391
Fax: 44-1784-49-5093
E-mail: packs@pg.com

OECD

Marie-Chantal Huet
OECD Environmental Health and Safety Division
2, rue André Pascal
75775 Paris Cedex 16
FRANCE

Tel: (33-1) 45-24-79-03
Fax: (33-1) 45-24-16-75
E-mail: marie-chantal.huet@oecd.org

Nicky Grandy
OECD Environmental Health and Safety Division
2, rue André Pascal
75775 Paris Cedex 16
FRANCE

Tel: (33-1) 45-24-16-76
Fax: (33-1) 45-24-16-75
E-mail: nicky.grandy@oecd.org

Herbert Koepp
OECD Environmental Health and Safety Division
2, rue André Pascal
75775 Paris Cedex 16
FRANCE

Tel: (33-1) 45-24-76-19
Fax: (33-1) 45-24-16-75
E-mail: herbert.koepp@oecd.org

ANNEX 2

Composition of working groups

	WG A	WG B	WG C
Chair	John Fenlon (UK)	Kees Romijn (GIFAP)	Michael Newman (US)
Rapporteurs	Helle Holst (DK) Leslie Touart (US)	Jacques Bedaux (NL) Dwayne Moore (CAN)	Colin Jansen (BE) Gerd Joermann (D)
Austria		Britte Grillitsch	Norbert Bornatowicz
Belgium	Katrin Delbecke	Isabelle Halleux	*[Colin Jansen, rapporteur]*
Canada	Peter Delorme	*[Dwayne Moore, rapporteur]*	Glen Atkinson
Denmark	*[Helle Holst, rapporteur]*	Niels Nyholm	Claus Hansen Gerard Jagers
Finland			Hannu Braunschweiler
France	Eric Vindimian	Jean-Francois Férard	
Germany	Martin Streloke	Sabine Martin Toni Ratte	
Italy		Silvia Marchini	
Netherlands	Rinus Bogers	*[Jacques Bedaux, rapporteur]*	Kees van Leeuwen
Norway	Erlend Spikkerud	Bjorn Dahl	
Spain	Enrique Moliner		
Sweden	Lars Lindqvist		
Switzerland	Roland Fisch		
UK	Mark Crane *[John Fenlon, chair]*	Andrew Riddle	Tim Sparks
USA	David Farrar	Richard Clements	Johanna Jaworska *[Michael Newman, chair]*
BIAC	Roland Maisch	Michael Harrass	Kathleen Stewart
ECETOC	Roger van Egmond	Lisa Tattersfield	
GIFAP		*[Kees Romijn, chair]*	
Steering Group	Bas Koiijman (NL) *[Leslie Touart (US), rapporteur]*	Simon Pack (UK) Jose Tarazona (ESP)	Peter Chapman (UK) Pascal Isnard (F) *[Gerd Joermann* (D), *rapporteur]* Reinhard Meister (D)

ANNEX 3

Report of Working Group A

Chairman: John Fenlon (UK)
Rapporteurs: Helle Holst (DK) and Leslie Touart (US)

SESSION 1: Why are tests performed and what do we want from the results?

Why do we have ecotoxicity tests?

The group focussed its discussion on aquatic ecotoxicity tests within the OECD framework. OECD is concerned with developing harmonized test guidelines generally for pesticides and industrial chemicals. These tests are used mainly to derive data for regulatory purposes (e.g. for classification and risk assessments within notification and registration schemes), in order to predict possible effects of the tested substance on the aquatic environment. It is worth noting, however, that other types of risk assessments use / require different testing procedures or guidelines (e.g., bioassays for monitoring or for site-specific assessments using contaminated sediments, soils or wastewater).

In consideration of these (mainly regulatory) purposes of OECD-type tests, the group recognized that:

⇒ standardization of the test design is essential for the mutual use of such data by different countries and for the use of the same data for both classification and risk assessment.

⇒ some flexibility in the designs is also necessary in order to cope with the vast variety of different substances and their physical-chemical properties.

With respect to possible changes of test designs, e.g. as a consequence of changes of the statistical analysis, the group also stated that:

⇒ OECD should then as well develop methods for the continued use of existing data. Repetition of tests should be avoided, where possible.

What kind of information do we need from ecotoxicity tests?

Information from laboratory testing is used to predict the possibility of effects in the real environment. This extrapolation involves a high degree of uncertainty from various causes. However, since field testing is not possible at the necessary scale and does not provide unambigious results, laboratory testing is necessarily being used for decision making. Therefore, it is generally agreed that

⇒ we must continue to improve testing and analytical techniques to move toward reducing uncertainty.

This involves two aspects: *more use of the generated data* and *better analysis*.

As to the *use made of the data*, the existing tests already generate a lot of information but frequently only one data point is ultimately expressed (e.g., NOEC). This may be sufficient in cases where only a relative measure of toxicity is needed (e.g., priority setting, classification, limit tests to identify substances of very low toxicity). Risk assessments (prediction of possible effects) obviously require better/absolute measures of toxicity and refined analysis. The group agreed that

⇒ results from tests providing more detail about effects other than NOEC approaches are more useful in risk assessments.

As to *improving the analysis of test results*, a major issue of reducing uncertainty is the introduction of a measure of precision of the test endpoint. Unlike classification schemes, risk assessments do consider a precision component. The group concluded that:

* New analytical techniques which provide for such a measure of precision and/or which reduce the extent of the necessary extrapolation are regarded to give better results, which is generally perceived as an improvement.
* The degree of accuracy needed for each endpoint and the use of this information in risk assessments needs to be discussed further.
* If new analytical techniques are to be introduced into the OECD test guidelines, the implications of the recommended technique(s) for existing test designs and for the continued use of existing data needs to be evaluated.

It should also be recognized that the perception of risk may vary, and, therefore, some countries may require more fixed criteria for using the endpoint in a risk assessment.

Are we happy with current practices? What should tests look like in future?

The group identified several major shortcomings of the current testing practice:

- **Waste of data/imprecise results**

* ANOVA-type determination of the NOEC (i.e., by comparing control and one treatment with hypthesis tests) does not use information from all the other treatment levels (i.e., the slope of the dose-response-curve).
* There is no measure of precision of the NOEC.
* The NOEC itself is imprecise because it can only be one of the test concentrations and because the power of the statistical tests frequently does not allow for detecting considerable effects (up to 20 % was mentioned). Thus, bad testing (high variability of controls) is rewarded, and the NOEC as one of the tested concentration is subject to decisions of the study director (the chosen concentrations and their spacing). This is regarded as scientifically inappropriate.

⇒ For all these reasons, the group concluded that the test design and/or the statistical analysis should be improved.

- **Ineffective use of test animals**

* By not using much of the data (see above), test animals are used inefficiently. This is not acceptable and needs to be improved.

- **Extrapolation needed**

* Both the NOEC and the ECx correspond to a standardized exposure time, species, and laboratory condition. Hence, an extrapolating factor is needed for their use in risk assessments. With mechanistic modelling, the need for part of this factor may be reduced (e.g., the part which accounts for the extrapolation from standardized to unlimited exposure time).

- **Lack of time component**

⇒ There was agreement that more information on the time component (effect build-up over time) and more use of such data is required. This was stated for both acute and (sub)chronic tests. Such dynamic information is an important aspect in risk assessments, e.g. for evaluating the probability and extent of effects (especially when using time-dependent fate data), for evaluating the relative risks for different effects (e.g., growth reduction versus reproduction impairment) and for risk/benefit analyses.

- **Further issues to consider:**

Perception of the NOEC: Most regulators do not use NOEC values without some further interpretation of the data. One approach is comparing the NOEC with an ECx, e.g. the EC10. An NOEC > EC10 should then be used only carefully and with low confidence. However, the NOEC is frequently misinterpreted as a true no-effect-level, especially with the lay public and with risk managers where there are more decision making levels. As to the US, participants stated that EPA is somewhat „happy" with the NOEC in the context of its regulatory use. In EPA evaluations, it usually represents low level effects which are not identified as statistically significant. While acknowledging the scientifc shortcomings and frequent confusion with an no-effect-level, the proper interpretation of other endpoints (ECx, NEC) by risk managers and the public would also be a point of concern.

Revised test design: In case of a move towards regression analysis, several issues would need to be addressed in detail:

* The optimized test design may need to be different for different organisms and endpoints of concern.
* The same applies to the value of "x" in ECx, due to different natural variability of endpoints like growth, reproduction, mortality with several organisms. The value of "x" and the chosen statistical technique also are likely to have implications on the number and spacing of test concentrations.
* Further, any change of the analytical technique should be accompanied by discussion and recommendations on the use of exisiting data in the future and on the parallel use of NOEC and the new measure(s) during a transition period.

SESSION 2: Review and comparison of the different approaches to data analysis

The group first focussed on the basic comparison between NOEC (hypothesis testing) and estimation procedures with regard to regulatory tests.

Should the NOEC still have a role in future testing or should we move away from it?

In reviewing this issue, the group collated the following views:

pro NOEC	contra
* easy to understand	* easy to misunderstand
* can always be determined	* is often misused or inappropriate
* big confidence intervals using regression	* no confidence intervals or other measureof precision with NOEC
* many regression models to choose between	* NOEC also can depend on the choice of model or statistical test
* untrue to say that only control/NOEC data are used in NOEC (ANOVA incorporates variance/df from all data, Williams' test considers additional dose related information)	* NOEC is not a sound and reasonable measure, for the reasons outlined in the background papers by S. Pack and P. Chapman and in the 1993 review report by S.Pack.

In conclusion, the group reached consensus on the following recommendation:

⇒ **OECD should move from the NOEC towards a regression-based (estimation) analysis of aquatic ecotoxicological data.**

The group also addressed briefly the general implications of a possible change towards regression or mechanistic modelling:

* Range finding tests for the optimal choice of test concentrations are necessary at any rate, regardless of the analytical technique.
* Test designs would need to be optimized for static regression analysis (e.g., more test concentrations, less replicates). Some modifications would also be needed for the dynamic approaches (time component), although existing data can already be used for modelling in DEBtox, for example. However, some changes would provide for better results.
* Guidance would need to be developed for the selection of the best model(s) (both static and dynamic) and for special cases (poor dose selection/ill-conditioned data, etc.) where model fitting fails.
* Procedures for the use of existing data sets would be necessary.
* The proper use of the additional information in risk assessments should be discussed.

Which other analytical technique could replace the NOEC ?

During an animated discussion about empirical versus mechanistic modelling, one view was that many empirical models were contrary to biological knowledge and frequently inconsistent. Another view, in the minority, was that an empirical, best-fit approach was better. It was further pointed out that empirical regression methods can also incorporate time-dependent hazard/survival data, and that an EC0 is model dependent and can be present (or not) in both empirical and mechanistic models. While doubts were expressed if a move to full-scale mechanistic modelling could be achieved in one step, the group recommended that:

⇒ **Dynamic (time-based) components should be incorporated into the regression models. The collection of time-course data should be extended.**

Which statistical measure(s) should be reported?

All raw data would be available to regulatory agencies. However, raw data normally are not used for decision making without analysis of some form, presently in determining an NOEC. For reasons of transparent and consistent decision-making by agencies and for planning purposes by industry, such 'reference point(s)' would be needed for risk assessment schemes in future. There was, however, a clear split of the group on whether NEC/EC0 values should serve this purpose (in a vote, 5 members expressed having problems with this concept while 9 had no such reservations, with 2 abstentions).

Without agreement on that issue, the group neverless concluded that:

* The form and parameters of the regression model should be reported together with confidence statements.
* A general model could probably be used for many datasets.
* If OECD decides to use mechanistic modelling, the EC0 should be reported.
* With the ECx - approach, different values of 'x' would probably be needed for different test procedures/endpoints, both in terms of matching up to the NOEC and of the achievable sensitivity. The value(s) of 'x' could either be chosen to correspond to the current level(s) of possible effects at the NOEC, or could be determined by using a sensitivity analysis of the optimized test design.
* For now, values of EC5 by increments of 5 to EC25 could be determined routinely.

Further, to facilitate the decision on which particular model(s) could be used in future, the group recommends that:

⇒ **OECD should undertake a study with existing (ring-test) data to compare different types of dynamic regression models.**

In conclusion, the group reached consensus on the following recommendations:

- ⇒ OECD should move from the NOEC towards a regression-based (estimation) analysis of aquatic ecotoxicological data.
- ⇒ Dynamic (time-based) components should be incorporated into these regression models. The collection of time-course data should be extended.
- ⇒ OECD should undertake a study with existing (ring-test) data to compare different types of dynamic regression models.

ANNEX 4

Report of the Working Group B

Chairman: Kees Romijn (FRA)
Rapporteurs: Jacques Bedaux (NL) and Dwaine Moore (CAN)

SESSION 1: Why are the tests performed and what do we want from the results ?

Why do we have ecotoxicity tests ?

Ecotoxicity testing can be conducted for several purposes, which can be grouped in different ways:

A) Prediction	Hazard identification Classification and priority setting Guidelines, Criteria Product development Risk assessment Research
B) Control and Monitoring	Permit compliance (monitoring/standards) Research
C) Diagnosis	Incidence reports Toxicity Identification and Evaluation (TIE) Research (identify causal agents)

or

A) Legal/Regulatory	New products Re-evaluations Site assessment Criteria, standards Permit setting and compliance
B) Non-regulatory	Waste management Product development and safety Market driven issues Emergency procedures Commercial issues Diagnosis
C) Research	Mode of action QSARs

Some of these applications of ecotoxicity data are beyond the scope of the workshop, which is to deal merely with substances and regulatory issues.

What kind of information do we need from these tests ?

The kinds of information we may wish to obtain from these tests are listed below. The types of information used in various applications is shown in table 1.

1) Lower levels of biological organisation:

biochemical
behaviour
survival
growth
reproduction

2) Population and higher levels of organisation:

Intrinsic rate of growth
Bioenergetic endpoints
Demographic endpoints
Richness, community structure
Community function endpoints

Data can be summarised using different parameters:

NOECs, LOECs, MATCs
EC_{50}, LC_{50}, IC_{50}
EC_x, LC_x, IC_p
LD_{50}
etc.

In the applications listed in table 1, various extrapolation procedures are available for use with the ecotoxicity data including:

Uncertainty factors
Models
Uncertainty analysis

The above are used to extrapolate from one species to another, from laboratory to field conditions, from short to long term effects, and from lower to higher biological levels of organization.

Are we happy with current practices ?

Participants agreed that the current summary parameters for ecotoxicity testing (e.g., NOEC) are inadequate and that other summary parameters and analytical techniques should be investigated.

SESSION 2: Review and comparison of the different approaches to data analysis

The topics discussed were:

(i) Should we use NOEC/ANOVA?
(ii) If not, what alternatives do we have?
(iii) What measures do we want to report?
(iv) How do we choose appropriate models?

Almost everybody agreed that ANOVA-type methods were not appropriate for estimating effective concentrations. However, if the aim is to test toxicity values against certain limits, ANOVA or other similar test procedures can be applied.

As a result, the following recommendations were made:

1) OECD tests should use a regression-based approach for the analysis of toxicity data.

2) For limit testing and other similar application, ANOVA-like methods can be applied.

Afterwards, the Group discussed what kind of regression models should be used. Two classifications were used: (1) static versus dynamic, and (2) empirical (or descriptive) versus theory-based (or mechanistic). Here static means that time is not incorporated in the model formulation. After a long discussion the following recommendations could be made:

3) OECD guidelines should encourage the characterization of the time-dependence when appropriate data are available (see table 2).

4) OECD should evaluate whether data should be collected at intermediate times for tests that do not currently have this requirement in the guideline.

Then, the Group discussed which summary measures are appropriate. This resulted in the following recommendations:

5) OECD should encourage the estimation of ECx,t including confidence limits for several values of x and, where appropriate, at different time intervals.

6) OECD should reanalyse existing ring test data to determine the precision associated with low ECx,t values for different biological test systems and several models. The objective would be to determine the lowest effect values that can be estimated with reasonable confidence for each test system.

7) Experimental designs for OECD test systems should be optimised for estimation of ECx.t for the last time interval.

8) OECD should determine:

(i) how dynamic data analyses would be used to effect better decision-making;
(ii) whether this improvement justifies the extra resources required to collect and analyze dynamic data for different tests.

Two issues were also identified in this session. The first issue is that the spacing of treatments for precise estimation of low ECx,t is difficult for all time intervals (see figure 1). Therefore, the Group recommended that:

9) OECD should not modify experimental test designs explicitly for dynamic analyses unless the benefits justify the additional costs of collecting and analyzing dynamic data.

The second issue raised concerned the choice of an appropriate low ECx,t value for risk assessment and other applications. Choosing such a value involves both statistical and regulatory considerations, and therefore was viewed to be outside the scope of this workshop. Nevertheless, various regulatory and other agencies should examine this issue carefully.

The next topic discussed was the procedure for choosing models. Group B did not agree with the initial Group C's proposal for choosing models based on the following priorities:

1. Best possible fit
2. Mechanistic
3. Empirical

An alternative proposal was made (which did not get consensus):

If models give equivalent and adequate fits and if assumptions are valid, mechanistic models are preferred over empirical models.

Concern was expressed about relying solely on best fit models because the chosen model could change between time intervals and between tests. A last general remark was that models should be as simple as possible and as complex as necessary.

Table 1: Commonly used summary parameters in different applications of ecotoxicity data.

Application	*NOEC LOEC*	*LC_{50} EC_{50}*	*LC_x EC_x*	*$LC_{x,t}$ $EC_{x,t}$*	*Population Level Measures*	*Community Level Measures*
Classification	xx	xx				
Criteria and Guidelines	xx	xx	xx			
Screening Level Risk Assessments	xx	xx	x			
Higher Tier Risk Assessments	x	x	xx	xx	xx	xx
Permit Compliance	x	xx	xx	??		
Product Development and Stewardship	xx	xx	xx	??		
Diagnosis (*e.g.*, TIE)	x	x	xx	x	x	
Research	x	x	xx	xx	xx	xx

x sometimes used xx frequently used

Table 2: Potentially useful summary parameters from an ecotoxicity test. It is also critical that the model equation, estimated parameters and their standard error, model goodness-of-fit, and other test results be reported as appropriate.

% Effect/Time	t_1	t_2	t_3	t_4
5	$EC_{5,t1}$±95% CI	$EC_{5,t2}$±95% CI	$EC_{5,t3}$±95% CI	$EC_{5,t4}$±95% CI
10	$EC_{10,t1}$±95% CI	$EC_{10,t2}$±95% CI	$EC_{10,t3}$±95% CI	$EC_{10,t4}$±95% CI
15	$EC_{15,t1}$±95% CI	$EC_{15,t2}$±95% CI	$EC_{15,t3}$±95% CI	$EC_{15,t4}$±95% CI
20	$EC_{20,t1}$±95% CI	$EC_{20,t2}$±95% CI	$EC_{20,t3}$±95% CI	$EC_{20,t4}$±95% CI
25	$EC_{25,t1}$±95% CI	$EC_{25,t2}$±95% CI	$EC_{25,t3}$±95% CI	$EC_{25,t4}$±95% CI
30	$EC_{30,t1}$±95% CI	$EC_{30,t2}$±95% CI	$EC_{30,t3}$±95% CI	$EC_{30,t4}$±95% CI
40	$EC_{40,t1}$±95% CI	$EC_{40,t2}$±95% CI	$EC_{40,t3}$±95% CI	$EC_{40,t4}$±95% CI
50	$EC_{50,t1}$±95% CI	$EC_{50,t2}$±95% CI	$EC_{50,t3}$±95% CI	$EC_{50,t4}$±95% CI

ANNEX 5

Report of Working Group C

Chairman: Michael Newman (US)
Rapporteurs: Colin Janssen (BE) and Gerd Joermann (D)

SUMMARY

Recommendations

The working group recommended that OECD should move away from the ANOVA/NOEC approach. and that future test methods should have the following qualities or improvements:

1. reduction of the possibility of biased results
2. more focus on biological significance of endpoints
3. more linkage to predictive, ecological or biological models allowing for the inclusion of relevant covariates where appropriate
4. inclusion of guidance for explicit statistical techniques in test guidelines
5. choice of summary statistics that can be generated and interpreted by non-statisticians.

To this end, the fitting of regression models was recommended: specific theory-based models, if appropriate, should be used for each individual test.

Future approach

The working group recommended test-specific regression models, theory-based if appropriate. It was agreed that test results should include the following:

1. estimates of all model parameters, error terms and goodness of fit.
2. a slope, if appropriate
3. ECx values
4. biologically relevant parameters
5. parameters describing the time course
6. confidence limits for summary statistics as is good statistical practice.

SESSION 1: Why are the tests performed and what do we want from the results?

Why do we have ecotoxicity tests?

The working group decided that the reason for conducting ecotoxicity tests were:

* to measure absolute or relative (ranking of) toxicity values
* for routine regulatory use
* to determine the need for and the design of higher tier tests
* to protect the environment by predicting the effects on natural populations (non human)
* to generate point estimates that are useful for assessment

What kind of information do we need from ecotoxicity tests?

The working group decided that the following type of information is needed:

* one or more endpoints (i.e. which is/are required for applying a classification scheme)
* time scale for effects
* biologically significant information
* indication and expression of reliability (accuracy, precision)
* information readily understandable by non statisticians
* representative of what is to be protected.

Are we happy with current statistical practices?

The working group identified the following shortcomings and problems in the currently used ANOVA/NOEC approach:

* unnecessarily high risk of false negatives
* no statements of biological significance, only statistical significance
* the type of data generated now will not meet future needs of risk assessment schemes
* inferior statistical methods
* no linkage to ecological, predictive models
* inadequate extration of information (e.g. only NOEC, no slope)
* NOECs lead to misunderstanding and misinterpretation
* detailed guidance on statistical methods is missing in guidelines
* no incorporation of covariates.

The working group recommended the following with regard to the design and performance of future tests and statistical analyses:

* Reduce the danger of false negatives and false positives.
* Develop tests that focus on biological significance.
* Make more efficient use of test information.
* Choose endpoints that are directly applicable to predictive ecological models and incorporate important covariates (e.g. time) where appropriate.
* The summary statistics/parameters chosen should be interpretable by non-experts.
* Make statistical tests an inherent part of test method development.
* Provide explicit recommendations for statistical data analysis in test methods and clear guidance on how to perform it/them. It must be possible for non-expert statisticians to perform the analyses.
* New statistical approaches should be compatible with GLP practice.
* Control variability should be taken into account in any analyses.
* A way to link new summary parameters to old endpoints is desirable.

SESSION 2: Review and comparison of the different approaches to data analysis

Should the NOEC/ANOVA approach be retained?

A large majority of working group participants did not want to retain the NOEC/ANOVA approach. However, one participant expressed some concern about replacing the NOEC for the following reasons:

- uncertainty with what it will be replaced with
- no infrastructure for moving away from NOEC (i.e. computer programs)
- NOEC has protected the environment in the past
- possible need to develop new safety factors (i.e. risk assessment practices)
- the current stringent requirements for design, doses, random variation lead to valid NOECs.

What could the NOEC/ANOVA approach be replaced with?

The following options were considered:

1. summary parameters for dose-response models which allow assessment (e.g. regression, mechanistic, mixed regression/ANOVA approach, others ..).
2. improve NOEC methods (e.g. test design, statistical test method, reporting, ...)
3. parallel reporting (both NOEC and ECx), including qualification of existing NOECs.
4. EPA interpolation method.

The working group agreed to concentrate further discussions predominately on option 1.

What type of information (endpoints; summary statistics) do we want?

- **Empirical vs Mechanistic**

There was no clear consensus on whether models should be empirical or mechanistic. Some participants did not feel comfortable with the mechanistic models (e.g. DEBtox) as they did not completely understand them. The pros and cons depended on the various factors (e.g. theory used, best fit, ...).

The working group did, however, propose the following logical sequence:

best (reasonable) fit -> mechanstic if possible -> if not, then use empirical models.

Most participants were in favour of theory-based models, fit by non linear regression.

- **Include covariates in the dose-response assessment?**

The working group agreed that models giving the best (reasonable) fit should be chosen first. Time should be included in this model, if appropriate. The inclusion of time *and* other covariates was selected as a second choice.

With respect to the inclusion of time as a covariate, the group acknowledge that for some existing tests, time could be included without changes to the method (e.g. *Daphnia* reproduction study). However, other tests would need to be redesigned and this caused some concern.

What type of summary statistics do we want from our selected approaches?

Finally, the working group identified the type of summary statistics that should be reported:

- estimates of all model parameters completed by terms of error, goodness of fit
- slope (with confidence limits)
- ECx (with confidence limits)
- biological parameters relevant for models
- parameters describing time course, if applicable.

Other summary statistics considered were:

- No Effect Concentration - NEC (or approximate of NEC)
- benchmark concentrations = lower confidence of ECx

Reservations were made with respect to the NEC. The fact that it is likely that a true NEC may exist for some chemicals and not for others gave concern as to its use as a parameter in chemical regulation.

ANNEX 6

OECD Workshop on Statistical Analysis of Data from Aquatic Ecotoxicology Tests

A Discussion of the NOEC/ANOVA Approach to Data Analysis

Dr. Simon Pack
Procter & Gamble, U.K.

- **Basic Principle**

  The NOEC/ANOVA approach is to compare each test concentration against the control.

  In general, the NOEC is the highest concentration that is not statistically significantly different from control. The LOEC is the lowest concentration that is significantly different from control.

  There may be some ambiguity in these definitions if, for a concentration significantly different from control, there is a higher concentration which is not significantly different. Some might prefer an alternative definition that the LOEC is the lowest concentration significantly different from control with the NOEC being the next lowest concentration.

- **Methodology**

  Start with parametric ANOVA as described in basic statistics textbooks and as implemented in all general statistical computer packages. Data should be transformed to satisfy the assumptions (the assumptions should always be checked to validate any analysis). The ANOVA assumptions are that the residuals are independently and identically normally distributed with zero mean and constant variance. With parametric ANOVA of aquatic toxicology data, insufficient attention is often paid to the mean-variance relationship of the data with the consequence that inappropriate estimates of variability are obtained leading to incorrect inferences. Generalised linear models do not seem to be used.

  Non-parametric versions of ANOVA exist that will generally be more robust without sacrificing much sensitivity. In fact, non-parametric methods could be used by default to avoid issues around violations of the assumptions.

  To compare test concentrations with the control there are many approaches (multiple comparison procedures). These take account of the multiple statistical tests by adjusting the statistical threshold for declaring significance. Dunnett's method is perhaps the most popular in ecotoxicology area. Williams' method uses the assumed trend in the underlying concentration-response.

  Despite the relative conceptual simplicity there are, in fact, a myriad of variants when the different combinations of analysis and multiple comparisons are taken into account.

- **Experimental Design**

  It is vital to identify what constitutes a replicate e.g. individual fish in a tank or the tank itself. Investigation of the components of variability may show whether grouping of individuals can be reasonably made.

The number of replicates will significantly impact the sensitivity of statistical analyses and therefore the NOEC. Increased control replication relative to each concentration will maximise sensitivity for a fixed overall number of replicates and is recommended. The control replication should be increased by a factor of $\sqrt{}$(no.of concentrations) relative to the test concentrations.

Optimising the design of ANOVA experiments from a statistical viewpoint seems to be seldom done.

- **Advantages of NOEC/ANOVA**

  - *Conceptual simplicity*

With ANOVA, we simply check if each concentration is different from the control. The modelling assumptions are relatively weak i.e. no concentration-response model is used. Distributional assumptions can largely be avoided by using non-parametric methods of analysis.

In comparison, regression methods aim to fit an empirical curve through the data points relating response to concentration. This curve need not be given any particular biological interpretation although one is often assumed. The main objective is to model the data. No-effect concentrations (NECs), hormesis and time can be readily incorporated. The adequacy of the model is assessed from the fit of the model to the data points.

Unlike ANOVA and regression methods, the mathematical modelling approach makes explicit assumptions about the underlying processes. These assumptions then generate a concentration-response model for the data. Unknown parameters are then estimated from the data. For some models the experimenter may need to supply values for fixed constants however these should still really be considered as parameters. The assumptions can only be verified by looking at the adequacy of the fit of the model to the data points.

  - *Computational simplicity*

Hand calculation is easy. Formulae are given in most basic statistics textbooks. Many specialist and non-specialist (e.g. spreadsheet) programs are available for ANOVA calculations.

In comparison, regression and mathematical modelling approaches require specialist software. Some simple (non-parametric) methods have been developed for the estimation of e.g. EC50s. However, these may not be sufficiently flexible for general recommendation. The more complex the assumptions and the model, the more complex the computational side will be and the more problems will be encountered.

  - *Experimental design is straightforward*

The number of replicates needed to give any degree of sensitivity can be readily calculated if the standard ANOVA assumptions are assumed.

Optimal experimental design for the regression and mathematical modelling approaches is possible but not straightforward and would most likely require further research.

- Disadvantages of NOEC/ANOVA

  - *The NOEC must be one of the test concentrations*

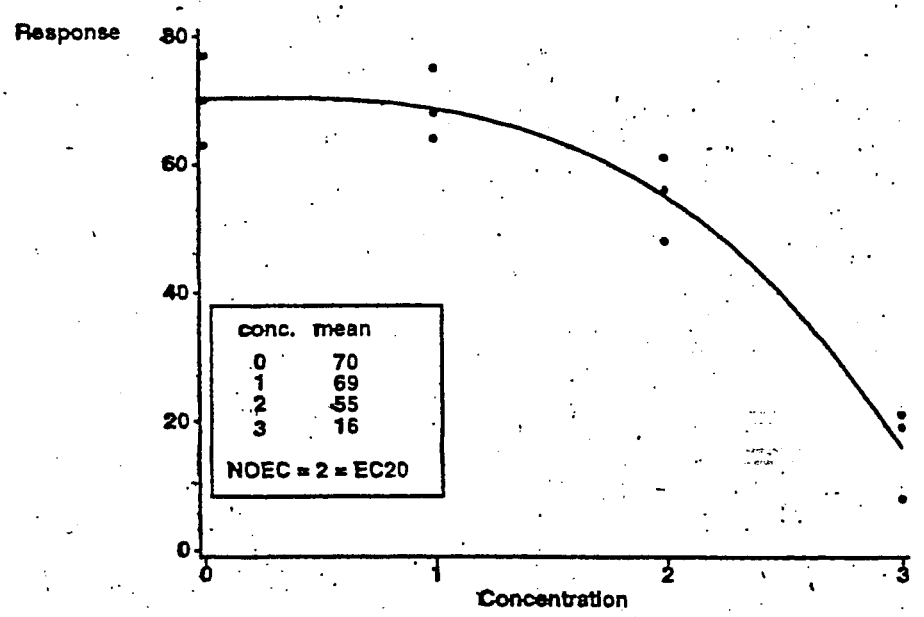
The NOEC is determined to a large extent by experimenter's choice of concentrations. Usually only a small number of concentrations are tested. Therefore the 'precision' is likely to be very limited.

  - *The NOEC is not a safe concentration*

If experimental variability is relatively high then the sensitivity of the analyses to detect differences from control will be relatively low. This implies that only larger differences from the control can be detected. This in turn implies the NOEC may be a concentration that actually corresponds to quite large effects.

Literature supports this in practice. For example, the results of the final *Daphnia magna* reproduction ring test showed sensitivity was such that effects up to 20% could have been declared as not significantly different from control. Effect sizes at the NOECs averaged around 10% but some corresponded to effects of 20-30%.

The example below (modified from a real experiment) illustrates the problem. There is clearly an effect at the NOEC which corresponds to an EC20.



NOEC obtained using Dunnett's method.
Fitted curve is 3-parameter logistic.

  - *It is impossible to derive an estimate of the precision for the NOEC*

A power calculation can quantify the sensitivity of the analyses in terms of the difference from control that could be reasonably detected. However, this doesn't address the real problem of how accurate the NOEC is. By definition the NOEC is just one of the concentrations. If decisions are to be made on environmental safety then it is surely important to quantify the degree of confidence.

In contrast, precision estimates are readily available from both the regression and mathematical modelling approaches.

- *There is no information on the concentration-response curve*

The rate of change of response with concentration may be useful in assessing how sensitive a species is. This information is not available from ANOVA/NOEC methods. Therefore valuable information is being wasted. Prediction of the effects at concentrations other than those studied is not possible.

Prediction of effects or effect concentrations is particularly simple with regression modelling and, arguably to a lesser extent, mathematical modelling approaches, since the explicit aim is to fit a model to the data points. If time is also incorporated then predictions of effects at a given time can also be readily made.

- *Robustness*

Parametric ANOVA is robust to moderate violations of assumptions (e.g. lack of normality). However, variations in the methodology may produce different NOEC values for the same data. The extent to which these NOECs might differ will largely depend on the spacing of the concentrations relative to the observed concentration-response. Not much seems to have been published on this.

For regression and mathematical modelling, the more complex the model used, the harder it is to validate against the data and robustness then becomes an issue.
The EC50 is a derived quantity and is known to be robustly estimated i.e. reasonably model-independent. More extreme percentiles, e.g. EC5s, will be highly model-dependent. However, estimates of precision will be correspondingly low, reflecting the information in the data, so that confidence intervals will generally overlap for different models. Therefore, in general, simplicity in the aims and approach is recommended.

- *The NOEC/LOEC may not exist*

If the lowest concentration tested produces a statistically significant difference from the control then the NOEC will not exist. If none of the concentrations is significantly different from control then the LOEC will not exist. To obtain a NOEC/LOEC the experiment would need to be re-run with different concentrations. The first experiment may therefore be considered to have been wasted. However, the experiment could still yield useful information on the concentration-response curve.

In contrast, both regression and mathematical modelling approaches may be able to derive useful quantities, for example EC50s, from such 'failed' experiments. The basic requirement would only be that the respective models can be fitted to the data.

- *Good experimental practice is not rewarded*

Generally, the poorer the experimental conduct the higher the variability in the data. This in turn means lower sensitivity to detect differences from control and consequently higher NOECs may result, falsely implying 'greater safety'. This is completely unacceptable.

Correctly applied regression and mathematical modelling approaches will, in general, produce estimates of the parameters of interest with precision that reflects the information and variability in the data.

- Other Issues

  - *NOEC/LOECs cannot be compared to ECx values*

  There have been attempts to correlate NOEC/LOECs with percentiles of the concentration-response curve e.g. EC20s. NOECs are fundamentally different from ECxs and as the NOEC is largely dependent on the experimental design then any correlations that have been found are almost entirely coincidental.

  - *NOECs are not NECs*

  The NOEC is not an estimate of the no-effect-concentration (NEC). As already explained, the NOEC can correspond to non-zero effects and does not estimate a 'safe' concentration. NOECs cannot be correlated with NECs or ECxs.

  - *Hybrid methods*

  Some authors have proposed methods aimed at overcoming the weaknesses of the NOEC. For example, Hoekstra and van Ewijk propose a two-stage procedure. Firstly, the highest concentration is found for which the effect is not estimated to be larger than 25% (the bounded-effect concentration). Then interpolation is used, from the end-point of the confidence interval for the effect size at this concentration, to estimate the concentration giving a 1% effect (or other small value). They argue that this calculation yields a conservative estimate of a concentration giving 'negligible' effects. While this may be the case, the concentration derived still suffers from the fact that an associated estimate of precision is not provided although one could presumably be derived. Existence of the bounded-effect concentration is also not guaranteed. However, their approach does go some way to limit the dependency of the 'safe' concentration on the experimental design.

References Relevant to the Debate Between the Relative Merits of NOECs and Effective Concentrations

Campbell, P.J. and S.P. Hoy. 1996. ED points and NOELs: how they are used by UK pesticide regulators. *Ecotoxicology* 5: (in press).

Chapman, P.F., M. Crane, J.A. Wiles, F. Noppert and E.C. McIndoe. 1996. Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5: 1-18.

Chapman, P.F., M. Crane, J.A. Wiles, F. Noppert and E.C. McIndoe (eds.). 1996. Asking the Right Questions: Ecotoxicology and Statistics. Report of a Workshop Held at Royal Holloway University of London, Surrey, U.K. SETAC-Europe.

Chapman, P.F. and P.M. Chapman. In press. A second warning: NOECs are inappropriate for regulatory use. *Environ. Toxicol. Chem.*

Chapman, P.M., R.S. Caldwell and P.F. Chapman. 1996. A warning: NOECs are inappropriate for regulatory use. *Environ. Toxicol. Chem.* **15**:77-79.

Cousens, R. and C.J. Marshall. 1987. Dangers in testing statistical hypotheses. Ann. Appl. Biol. 111:469-476.

Dhaliwal, B.S., R.J. Dolan and R.W. Smith. 1995. A proposed method of improving whole effluent toxicity data interpretation in regulatory compliance. *Water Environ. Res.* **67**: 953-963.

Dhaliwal, B.S., R.J. Dolan, C.W.Batts, J.M. Kelly and R.W. Smith. 1996. Warning: Replacing NOECs with point estimates may not solve regulatory contradictions. A fundamental change in toxicity data interpretation is warranted. *Envion. Toxicol. Chem.* In press.

Hoekstra, J.A. and P.H. Van Ewijk. 1993. Alternatives for the no-observed-effect level. *Environ. Toxicol. Chem.* **12**:187-194.

Kooijman, S.A.L.M. 1996. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* **75**: 310-316.

Lacey, R.F. and M.J. Mallett. 1991. Further statistical analysis of the EEC ring test of a method for determining the effects of chemicals on the growth-rate of fish. Unpublished report of an OECD meeting of experts on aquatic toxicology, WRc, Medmenham, UK, 10-12 December, 1991.

Laskowski, R. 1995. Some good reasons to ban the use of the NOEC, LOEC and related concepts in ecotoxicology. *Oikos* **73**: 140-144.

Leisenring, W. and L.M. Ryan. 1992. Statistical properties of the NOAEL. *Regul. Toxicol. Pharmacol.* 8:161-171.

Masters, J.A., M.A. Lewis, D.H. Davidson and R.D. Bruce. 1991. Validation of a four-day ceriodaphnia toxicity test and statistical considerations in data analysis. *Environ. Toxicol. Chem.* **10**:47-55.

Moore, D.R.J. and P. Caux. 1996. Replacong NOELS AND LOELS with ECX point estimates as the prefered measure of low toxic effect. Environ. Toxicol. Chem. (in press).

Noppert, F., N. Van der Hoeven and A. Leopold. 1994. How to measure no effect : towards a new measure of chronic toxicity in ecotoxicology. Workshop Report of the Netherlands Working Group on Ecotoxicology. The Hague

Pack, S. 1993. A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines. Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent, ME9 8AG, UK.

Petersen, R.G. 1977. Use and misuse of multiple comparison procedures. Agronomy Journal 69:205-208.

Skalski, J.R. 1981. Statistical inconsistencies in the use of no-observed-effect levels in toxicity testing. In D.R. Branson and K.L. Dickson, Eds., Aquatic Toxicology and Hazard Evaluation, Fourth Conference, ASTM STP 737, American Society for Testing and Materials.

Stephan, C.E. and J.W. Rogers. 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. In R.C. Bahner and D.J. Hansen, Eds., Aquatic Toxicology and Hazard Evaluation, Eighth Symposium, ASTM STP 891, American Society for Testing and Materials.

Suter, G.W. II. 1996. Abuse of hypothesis testing statistics in ecological risk assessment. *Human Ecol. Risk Asssess.* **2:** 331-347.

References on Threshold and Hormesis Models

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

Cox, C. 1987. Threshold dose-response models in toxicology. Biometrics 43:511-523

Ewijk, P.H. van and Hoekstra, J.A. 1993. Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. Ecotoxicology and Environmental Safety. 25:25-32.

Yanagimoto, T. and E. Yamamoto. 1979. Estimation of safe doses : critical review of the hockey stick regression method. Environmental Health Perspectives 32:193-199.

References on the Benchmark Concentration

Crump, K.S. 1984. A new method for determining allowable daily intakes. Fundam. Appl. Toxicol. 4:854-871.

Crump, K., B. Allen and E. Faustman. 1995. The use of the benchmark dose approach in health risk assessment. Report of the US EPA Risk Assessment Forum, EPA/630/R-94/007, February 1995.

References on Time to Response Modelling

Dixon, P.M. and M.C. Newman. 1991. Analysing toxicity data using statistical models for time-t0-death: An introduction. In: Metal Ecotoxicology, Concepts and Applications. Edited by Michael C. Newman and Alan W. McIntosh. Lewis Publishers, Inc., Chelsea, Michigan.

Kooijman, S.A.L.M.. 1993. Dynamic Energy Budgets in Biological Systems: Theory and Applications in Ecotoxicology. Cambridge University Press, Cambridge.

ANNEX 7

OECD Workshop on Analysis of Data From Aquatic Toxicology Tests

Alternatives to the NOEC Based On Regression Analysis

Peter Chapman
Zeneca Agrochemicals

Effective Concentration (EC) Estimation

An "effective concentration" (EC) is defined as the concentration that produces a specified size of effect relative to an untreated control. An equivalent definition applies to an effective dose, but concentration is used throughout this document.

Methodology

A typical experiment comprises an untreated control plus a number of concentrations replicated a number of times. The measurement made on each experimental unit is some form of sub-lethal response such as the weight of an organism or the number of offspring produced. A regression model is fitted to the data and, through a process known as inverse estimation, a concentration corresponding to a specified percent effect relative to the control is estimated. For example, a concentration corresponding to a 50% reduction in effect relative to the control could be estimated. Figure 1 illustrates how an EC is estimated for a typical sigmoidal dose-response curve. Confidence intervals for the EC can, and should always be, estimated.

The curve fitted will usually be empirical in nature and will not have any particular biological justification. However, experience demonstrates that many curves, such as that illustrated in figure 1, fit ecotoxicity data well and are adequate for the purpose of estimating effective concentrations.

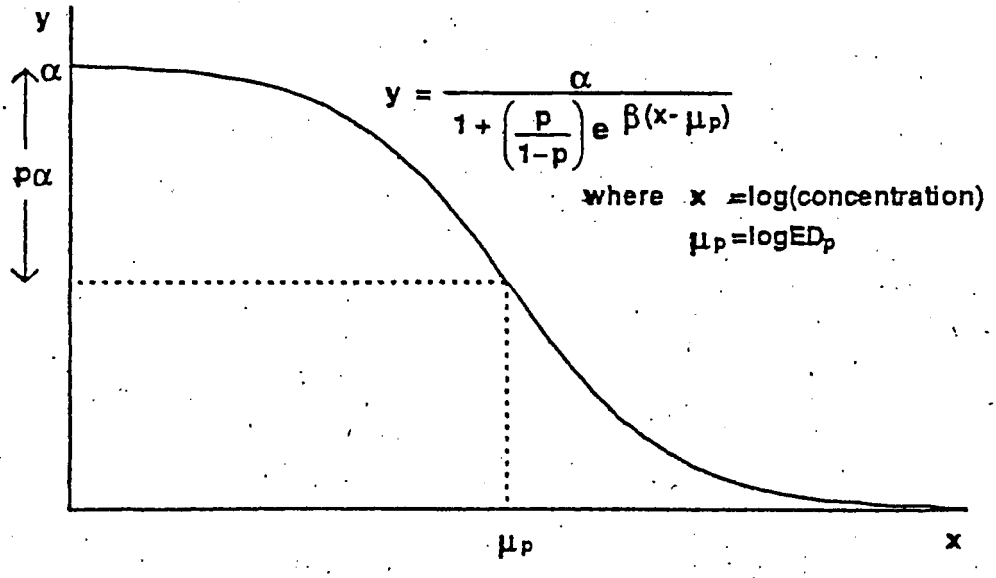


Figure 1 : Concentration-Response curve for the logistic regression equation.

Experimental Design

The design will usually take the form of a fully randomised or randomised complete block, although more complicated designs are possible. If it has been decided to replicate the control and the doses, it is important that replication is true replication and not pseudo-replication. (Pseudo-replication occurs when sub-samples from an experimental unit, such as measurements on individual organisms within a single housing unit, are used in the analysis of an experiment as if they are true replicates.) The number of replicates need not be the same for each dose.

Potentially Useful Extensions To Regression Procedures

Benchmark Concentration

The benchmark dose (BC) is defined as the statistical lower confidence limit on a concentration which produces some pre-determined increase in response rate compared to the untreated control. In other words, it is the lower confidence limit on an EC estimate. The BC is a relatively conservative quantity on which to base the estimate of a safe concentration, but this characteristic has found it many advocates in fields such as mammalian toxicology where the potential risk to humans must be kept very small indeed. It may however be considered too conservative for use in ecotoxicology.

The EC_0

Commonly used dose-response models, such as the logistic curve, predict non-zero effects even at very small doses so an EC estimate cannot be regarded as an estimate of a true No Effect Concentration (NEC). It is possible, however, to modify conventional dose-response models in order to estimate an EC_0. Two such types of model which deserve serious consideration are threshold models and hormesis models.

Threshold Models

Threshold models are based upon the supposition that there exists a dose at, and below which, a substance produces no toxic effect and above-which an increasing effect occurs. Threshold models can be thought of as two models joined together at a point. One part is a horizontal line describing the level of background - or untreated - response; the other describes an increasing effect with increasing dose. The two parts join at the EC_0 or threshold dose (see figure 2). This is the maximum dose at which the response is equal to the background and can be included in the model specification as a parameter and so be estimated directly.

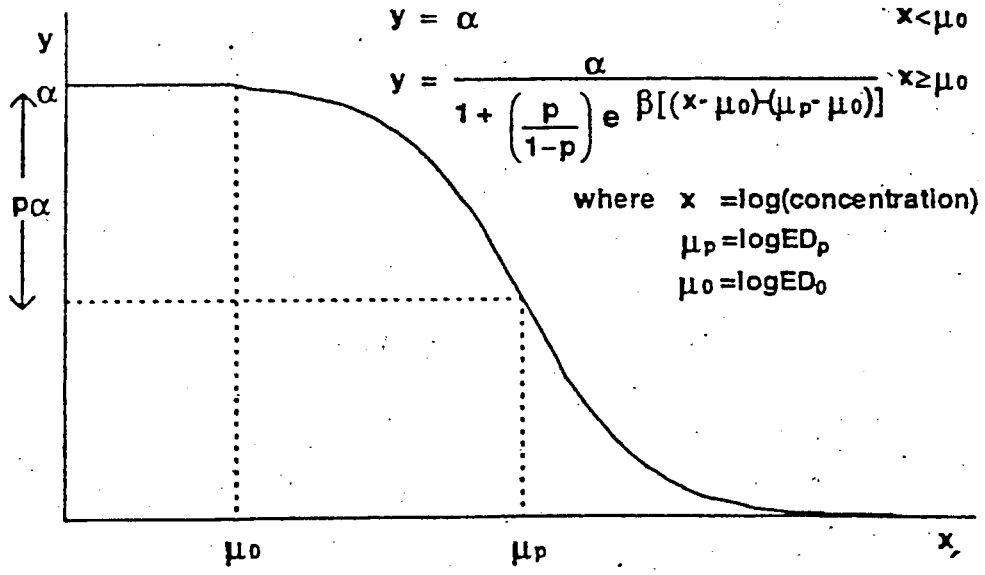


Figure 2 : Calculation of a threshold concentration or EC_0.

Hormesis Models

Regression models fitted to dose-response data are generally monotonic, reflecting an ever-increasing adverse effect with increasing dose. Problems arise, however, when we come across effects which seem to contradict this expected monotonicity. A particular example of this is hormesis, in which low doses of a substance appear to stimulate an apparently beneficial response in the test organism even though larger concentrations lead to a toxic effect (see figure 3). From an ecotoxicological perspective, the point at which the response is equal to the untreated response - the EC_0 in figure 3 - could be interpreted as the concentration at which the stimulatory effect ceases and the toxic effect begins and thus could be regarded as an estimate of the true NEC.

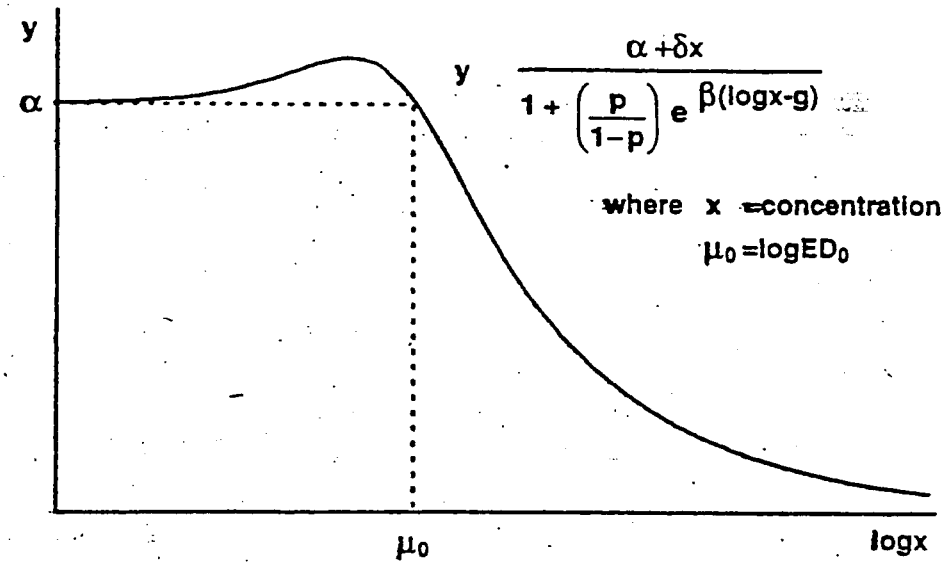


Figure 3 : EC_0 estimation in the presence of stimulation of response at low concentrations (hormesis).

Time to Response Models

If the responses in a single ecotoxicity test are measured on a number of different days, then a sigmoidal concentration response curve can be modified to include time. One of the benefits is that EC estimates can be given for different times. An additional benefit is that time to response can be estimated as a function of concentration.

Advantages and Disadvantages of EC Estimates

Below is a list of advantages and disadvantages of EC estimation as an alternative to the NOEC. The list of advantages mainly draws attention to ways in which EC estimates overcome the scientific deficiencies in the NOEC. In contrast the list of disadvantages describes some of the difficulties to be encountered in trying to estimate EC points. All of these difficulties can be overcome, although to do so may require very large experiments or a significant amount of resource in analysing data from an experiment.

Advantages

1 Regression permits the estimation of effects at untested doses. In contrast, a NOEC can only be one of the doses actually used in the experiment.

2 EC estimation rewards well conducted experiments - i.e those which are unbiased and of low variability. The greater the precision in an experiment the smaller will be the width of the confidence interval around the estimate. This implies that the lower the variability in an experiment the the more likely it is that an EC estimate will be close to its true value. In contrast, highly variable results are more likely to produce EC estimates that are much larger, or much lower, than their true values.

Standard ANOVA/NOEC evaluation, on the other hand, rewards poor experiments (i.e. high variability) with high NOEC values.

3 Because confidence intervals can always be calculated, the precision of an EC estimate can always be determined, whereas the precision on a NOEC can never be determined.

4 An EC can always be estimated, even if it is outside the range of experimental doses, although estimates far outside the range of doses should be treated with caution.

By contrast, the NOEC is not always obtainable, either because the lowest dose gives a statistically significant effect or none of the doses do. This situation has resulted in tests having to be repeated.

5 The biological effect produced by a dose equal to the EC is not zero but it is known because it is pre-selected. This situation is to be preferred to that of the NOEC, which also produces a non-zero effect which can be quite large, but is of unknown magnitude.

6 Unlike the NOEC, the estimated EC does not depend upon the type I error rate in a significance test, nor on the choice of multiple comparison procedure.

Disadvantages

1 The choice of regression model: If one is interested in estimating an EC for a small effect size, such as an EC10 or EC5, the estimate will usually depend upon the choice of model. For any single set of test results, therefore, the data analyst may have to fit a large number of models, and even then it may be difficult to decide which is best.

2 Even if one is confident that the right model has been fitted, for small effect sizes the confidence interval around the EC estimate will be relatively large.

3 Precision of EC estimates depends upon the number of test concentrations and their values. Therefore, if it is important to be able to estimate small effect concentrations, some research will need to be carried out to determine optimum selection of test concentrations. This may need to be done separately for each model.

4 Choice of effect size: If EC estimation is to replace the NOEC, then for each type of endpoint in each type of test the size of effect of interest needs to be decided upon.

5 For Benchmark Concentrations not only do the model and the effect size need to be chosen but also the confidence level used in estimating the confidence interval.

6 Both threshold and hormesis models require at least one extra parameter compared with a normal sigmoid model. This makes it more difficult to fit the model.

7 Experience suggests that confidence intervals around NEC estimates from threshold and hormesis models tend to be very wide. Thus before these models can be used some research into the optimal selection and placing of doses is required.

References Relevant to the Debate Between the Relative Merits of NOECs and Effective Concentrations

Campbell, P.J. and S.P. Hoy. 1996. ED points and NOELs: how they are used by UK pesticide regulators. *Ecotoxicology* 5: (in press).

Chapman, P.F., M. Crane, J.A. Wiles, F. Noppert and E.C. McIndoe. 1996. Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5: 1-18.

Chapman, P.F., M. Crane, J.A. Wiles, F. Noppert and E.C. McIndoe (eds.). 1996. Asking the Right Questions: Ecotoxicology and Statistics. Report of a Workshop Held at Royal Holloway University of London, Surrey, U.K. SETAC-Europe.

Chapman, P.F. and P.M. Chapman. In press. A second warning: NOECs are innappropriate for regulatory use. *Environ. Toxicol. Chem.*

Chapman, P.M., R.S. Caldwell and P.F. Chapman. 1996. A warning: NOECs are innappropriate for regulatory use. *Environ. Toxicol. Chem.* 15:77-79.

Cousens, R. and C.J. Marshall. 1987. Dangers in testing statistical hypotheses. Ann. Appl. Biol. 111:469-476.

Dhaliwal, B.S., R.J. Dolan and R.W. Smith. 1995. A proposed method of improving whole effluent toxicity data interpretation in regulatory compliance. *Water Environ. Res.* 67: 953-963.

Dhaliwal, B.S., R.J. Dolan, C.W.Batts, J.M. Kelly and R.W. Smith. 1996. Warning: Replacing NOECs with point estimates may not solve regulatory contradictions. A fundamental change in toxicity data interpretation is warranted. *Envion. Toxicol. Chem.* In press.

Hoekstra, J.A. and P.H. Van Ewijk. 1993. Alternatives for the no-observed-effect level. *Environ. Toxicol. Chem.* 12:187-194.

Kooijman, S.A.L.M. 1996. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* 75: 310-316.

Lacey, R.F. and M.J. Mallett. 1991. Further statistical analysis of the EEC ring test of a method for determining the effects of chemicals on the growth-rate of fish. Unpublished report of an OECD meeting of experts on aquatic toxicology, WRc, Medmenham, UK, 10-12 December, 1991.

Laskowski, R. 1995. Some good reasons to ban the use of the NOEC, LOEC and related concepts in ecotoxicology. *Oikos* 73: 140-144.

Leisenring, W. and L.M. Ryan. 1992. Statistical properties of the NOAEL. *Regul. Toxicol. Pharmacol.* 8:161-171.

Masters, J.A., M.A. Lewis, D.H. Davidson and R.D. Bruce. 1991. Validation of a four-day ceriodaphnia toxicity test and statistical considerations in data analysis. *Environ. Toxicol. Chem.* 10:47-55.

Moore, D.R.J. and P. Caux. 1996. Replacong NOELS AND LOELS with ECX point estimates as the prefered measure of low toxic effect. Environ. Toxicol. Chem. (in press).

Noppert, F., N. Van der Hoeven and A. Leopold. 1994. How to measure no effect : towards a new measure of chronic toxicity in ecotoxicology. Workshop Report of the Netherlands Working Group on Ecotoxicology. The Hague

Pack, S. 1993. A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines. Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent, ME9 8AG, UK.

Petersen, R.G. 1977. Use and misuse of multiple comparison procedures. Agronomy Journal 69:205-208.

Skalski, J.R. 1981. Statistical inconsistencies in the use of no-observed-effect levels in toxicity testing. In D.R. Branson and K.L. Dickson, Eds., Aquatic Toxicology and Hazard Evaluation, Fourth Conference, ASTM STP 737, American Society for Testing and Materials.

Stephan, C.E. and J.W. Rogers. 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. In R.C. Bahner and D.J. Hansen, Eds., Aquatic Toxicology and Hazard Evaluation, Eighth Symposium, ASTM STP 891, American Society for Testing and Materials.

Suter, G.W. II. 1996. Abuse of hypothesis testing statistics in ecological risk assessment. *Human Ecol. Risk Asssess.* 2: 331-347.

References on Threshold and Hormesis Models

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

Cox, C. 1987. Threshold dose-response models in toxicology. Biometrics 43:511-523

Ewijk, P.H. van and Hoekstra, J.A. 1993. Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. Ecotoxicology and Environmental Safety. 25:25-32.

Yanagimoto, T. and E. Yamamoto. 1979. Estimation of safe doses : critical review of the hockey stick regression method. Environmental Health Perspectives 32:193-199.

References on the Benchmark Concentration

Crump, K.S. 1984. A new method for determining allowable daily intakes. Fundam. Appl. Toxicol. 4:854-871.

Crump, K., B. Allen and E. Faustman. 1995. The use of the benchmark dose approach in health risk assessment. Report of the US EPA Risk Assessment Forum, EPA/630/R-94/007, February 1995.

References on Time to Response Modelling

Dixon, P.M. and M.C. Newman. 1991. Analysing toxicity data using statistical models for time-t0-death: An introduction. In: Metal Ecotoxicology, Concepts and Applications. Edited by Michael C. Newman and Alan W. McIntosh. Lewis Publishers, Inc., Chelsea, Michigan.

Kooijman, S.A.L.M.. 1993. Dynamic Energy Budgets in Biological Systems: Theory and Applications in Ecotoxicology. Cambridge University Press, Cambridge.

ANNEX 8

Dynamic measures for ecotoxicity *

S.A.L.M. Kooijman, J.J.M. Bedaux
Vrije Universiteit, Dept. Theoretical Biology, de Boelelaan 1087, NL-1081 HV Amsterdam
A.A.M. Gerritsen, H. Oldersma & A.O. Hanstveit
TNO-Toxicology, Dept. Envir. Toxicology, P.O.Box 6011, NL-2600 JA Delft

12 September 1996

Abstract

There are three required components of dynamic models for toxic effects: toxico kinetics, effects on a target parameter coupled to the internal concentration and the physiological component. The Dynamic Energy Budget (DEB) model, which is used to model the latter component, relates a change in a target parameter of a particular physiological process, such as the specific costs for growth, to an output variable, such as the cumulative number of offspring. We compare the logit/probit and the DEB-based models conceptually and numerically and conclude that the DEB-based model is more effective as an effect model. The DEB-based model solves the problem of estimating the No-Effect Concentration and provides the required information to evaluate the consequences of effects on individuals for population dynamics.

1 Introduction

In environmental risk assessments Predicted No-Effect Concentrations (PNECs), that are derived from No-Observed Effect Concentrations (NOECs) in standard single species toxicity tests, are compared with Predicted Environmental Concentrations (PECs), derived from production volumes, use patterns, and transport in the environment. The purpose of the toxicity tests, together with their designs and experimental protocols, evolved gradually towards sophistication. The analysis of the results of these tests, however, did not catch up with these changes for a long time. This document introduces a process-based approach for the analysis of toxicity tests, which has a firm rooting in biology.

The purpose of this document is to identify the aim of toxicity tests and to present the DEB-based effect model as a method to achieve this aim. The static and the dynamic methods are compared, conceptually and numerically. We tried to refrain from technical details in this paper and refer to Kooijman and Bedaux (1996) for a full discussion of the

*Background Document for the OECD Workshop on Analysis of Data from Aquatic Ecotoxicity Tests, Braunschweig, Oct 1996

DEB-based model for toxic effects. A short introduction to the DEB-theory is given in a separate document.

1.1 Aims of toxicity tests

The primary purpose of standard toxicity tests is to provide information about

- the maximum concentration that gives no effect on a response variable (such as survival, body growth, reproduction, population growth). This information is used to derive a level in the environment that can be considered as "safe".
- what effects are to be expected in the environment if these levels are exceeded (a little bit)?
- the requirement for further in-depth studies with respect to the ecotoxicity of the chemical. The priority is high if expected levels in the environment are likely to approach or exceed the level that is considered "safe" and, in that case, the effect is likely to be substantial. The actual decision to further research depends on financial possibilities for research and socio-economic factors in industrial activity.

The second and third application of toxicity tests imply a quantification of effects as a function of the concentration. They also imply an extrapolation of the effects as observed during the tests to effects at long-term exposure and a translation to effects in the environment.

1.2 Classification of approaches

The analyses of toxicity tests can be classified into three approaches, the ANOVA method, static methods and dynamic methods.

- The ANOVA method aims to identify the highest tested concentration with a response that does not deviate from the blank on the basis of a statistical procedure: the NOEC.
- The static method quantifies effects at a standardized exposure time on the basis of concentration response models, such as the logit, probit or Weibull model, which are all very similar. The toxic effect is quantified by EC50 (LC50) or an ECx value for some small value of x. The method is called static because information about the rate at which effects build up during exposure is not used. Fixed extrapolation factors are used to predict effects at long exposure.
- The dynamic method quantifies effects as functions of the concentration *and* the exposure time. The Dynamic Energy Budget (DEB)-based model is an example of this approach. The toxic effect is quantified by two parameters: the no-effect concentration (NEC), and the killing rate (for survival) or the tolerance concentration (for other endpoints) as a measure of the effect if the NEC is exceeded.

2 The dynamic approach

The dynamic approach for the analysis of standard toxicity tests characterizes toxic effects with a No-Effect Concentration (NEC), a tolerance concentration (or a killing rate in the case of effects on survival), and an elimination rate. These parameters directly relate to long-term exposure, so no extrapolation factor is required here. They relate to changes in processes. From these three parameters, it is possible to calculate the static parameters, including the EC0 at the end of the test (therefore comparable to the NOEC), as well as the full ECx-time behaviour. It is not possible to calculate the dynamic parameters from the static ones, which shows that the dynamic parameters contain more information than the static ones. The dynamic approach uses information about the rate at which effects build up during exposure. It is important to realize that static and dynamic approaches are alternative analyses for the *same* toxicity tests. The approaches only differ in the type of information that is extracted from the data. Given the choice for the analysis of the data, we can try to optimize the experimental setup of the test in terms of efficiency. This is another issue that is beyond the scope of this paper.

2.1 The components of dynamic effect models

The three components that any dynamic model for effect should have are the kinetics, the effect and the physiological component, which are discussed in the next subsections. The implication being that responses are modelled as function of the concentration *and* exposure time. We have a response *surface*, rather than a concentration response relationship. For a response such as the cumulative number of offspring, this means that we include any delay of the start of the reproduction into the analysis.

2.1.1 Kinetics

The kinetics component links internal concentrations to external concentrations. The first order kinetics (also called the one-compartment model) is the simplest choice. It assumes that uptake is proportional to the external concentration and elimination is proportional to the internal concentration. If the organism grows during exposure, dilution by growth results in a deviation from first order kinetics, that should be taken into account. The DEB-based model takes uptake and elimination rates proportional to the surface area, while the DEB component (see below) specifies the growth process.

Actual measurements of time profiles for the internal concentration sometimes show deviations from first order kinetics. More-compartment models are frequently used to improve the fit, because they have more parameters. A more detailed modelling of the various uptake rates (via water and/or food), elimination rates (via water, reproduction and/or respiration), changes in fat content, and metabolic transformations, is frequently more realistic than using multi-compartment models. Some chemicals are not taken up at all, but have their effect on the outer side of the organism. Multi-compartment models should only be used if the compartments are identified and their concentrations of toxicant measured. The latter requirement applies to all kinetics models that are more complex

than the first-order one. Reconstructions of complex kinetics from observed effects, rather than from measured internal concentrations, run into problems rather easily.

Since internal concentrations are not measured in standard toxicity tests, application of more advanced models for kinetics in these toxicity tests is not feasible. The alternative to refrain from dynamic modelling and apply an arbitrary extrapolation factor to arrive at predicted long-term effects is certainly not a better alternative. We consider the application of a safety factor to compensate for possibly inappropriate use of first-order kinetics to be a more promising alternative. This problem is inherent to the use of standard toxicity tests for environmental risk assessments; an in-depth scientific study for long-term effects is the only sensible alternative. Because of the financial costs of such a study, this will only be possible for a limited number of chemical compounds.

2.1.2 Effects

The effect component links effects on a target parameter, such as the volume-specific costs for maintenance or growth, to the internal concentration. The linear effect model is the simplest choice: The change in the target parameter is proportional to the internal concentration that exceeds the internal No Effect Concentration (NEC). Translation of this concept into formulae gives the following relationships for sublethal and lethal effects on a target parameter: if int.conc(t) $\geq$ int.NEC at exposure time t

$$\text{par}_c(t) = \text{par}_0 \left(1 + \frac{\text{int.conc}(t) - \text{int.NEC}}{\text{int.tolerance conc.}}\right)$$

$$\text{haz}_c(t) = \text{haz}_0 + \text{killing rate}\,\frac{\text{int.conc}(t) - \text{int.NEC}}{\text{BCF}}$$

where BCF stands for the BioConcentration Factor (the ratio of the internal and the external concentration after long-term exposure) and haz stands for the hazard rate, i.e. the instantaneous death rate. As long as the internal concentration is less than the internal NEC, we have that par_c=par_0 and haz_c=haz_0. If we divide the internal concentrations by BCF, we get external concentrations. The result is for conc(t) $\geq$ NEC

$$\text{par}_c(t) = \text{par}_0 \left(1 + \frac{\text{conc}(t) - \text{NEC}}{\text{tolerance conc.}}\right)$$

$$\text{haz}_c(t) = \text{haz}_0 + \text{killing rate}\,(\text{conc}(t) - \text{NEC})$$

where conc(t) has the dimensions of an external concentration but it is proportional to the internal one. Note that the NEC refers to long-term exposure. The NOEC for a particular toxicity test depends on the maximum observed exposure time t. It is therefore conceptually more or less comparable with the EC0.t, while the NEC equals the EC0.∞.

The above mentioned formulae define the killing rate and the tolerance concentration; these parameters occur as proportionality coefficients in the description of effects. The killing rate has dimension 'per environmental concentration per time' and is a measure for the toxicity of a chemical if it exceeds the NEC; a toxicity measure that is independent of

the exposure time. The tolerance concentration has the dimension 'environmental concentration'. The more toxic the chemical, the lower is its value. It is essential to specify the target parameter to which the tolerance concentration relates.

Five target parameters are distinguished for effects on reproduction: two for direct and three for indirect effects. Direct effects on reproduction are defined as effects on what happens with the investment into reproduction, not on the size of investment itself. One mode of action increases the energetic costs of each young, the other affects the survival of each ovum during a short sensitive period. Indirect effects on reproduction affect the investment into reproduction, not the conversion of this investment into young. Indirect effects increase the costs of growth or maintenance, or decrease the assimilative input. These three indirect effects not only reduce the reproduction, but also delay the start of the reproduction.

Consistent with the five target parameters for effects on reproduction, there are three target parameters for effects on body growth: one direct effect (the increase of the specific costs for growth), and two indirect ones: the increase of maintenance costs and the decrease of the assimilative input.

For effects on population growth of algae, we distinguish a direct effect on cell growth (by increasing the specific costs for growth), and two effects on survival (so the target parameter is the hazard rate): one effect lasts during population growth and the other only operates during a very short period after inoculation. In the latter case, effects are supposed to occur only during the transition from the culture to the experimental conditions. The sensitivity of the cells is here supposed to relate to the cell cycle. The overall effect is a delay of population growth, rather than a reduction. The uptake/elimination kinetics is assumed to be fast relative to the population growth process, so that the internal concentration equals the product of the BCF and the external concentration.

The linear relationships between the internal concentration and the target parameter follow from two arguments: effective molecules operate independently, and it is a simple approximation for small changes in the target parameter. The improvement of goodness of fit for large changes in the target parameter probably does not balance extra parameters. Moreover, at higher concentrations, more physiological processes are likely to be affected simultaneously. This makes that many parameters have to be introduced to capture large effects. We consider such approvements counter productive in standard toxicity tests. The argument implies that if the goodness of fit of the model to the responses is less than excellent, a higher weight should be given to responses at low concentrations.

Note that a linear relationship between the target parameter and the internal concentration, does not imply a linear relationship between the response (i.e. output variable) and the external concentration. These relationships work out to be sigmoid.

2.1.3 Blank physiology

For acute lethal effects we do not need a model for the blank physiology. For effects on algal population, such a model can be very simple. For effects on body growth and reproduction, however, we need a physiological component that links output variable(s) to the target parameter. The output variable is the variable that is measured, such as body

length, cumulative number of offspring, number of surviving individuals, etc. The Dynamic Energy Budget (DEB) model is the simplest choice that links all essential processes: feeding, digestion, respiration, maintenance, growth, development, reproduction and aging. It is introduced in a separate document. Many other models for the uptake and use of food have been described in literature, which typically involve many parameters. A comparative discussion of these models is beyond the scope of this paper. Since the physiological component only relates to the ecophysiology of the test species that is involved, and not to the toxicant, its applicability needs not to be studied for each individual toxicity test. It needs to be studied only once per species. The physiological component in this respect differs fundamentally from the other two components: the kinetics and the effect component.

Some chemicals, such as endocrine disrupters, might have an effect at the molecular level that does not directly relate to the energetics of the organism, but other effects will eventually translate into effects on the energetics; it is the choice for output variable (e.g. growth or reproduction) that directly relates to the energetics, not the molecular mode of action of the chemical. If a chemical has neither direct nor indirect effects on energetics, such as a chemicals that only affects behaviour, it will have no effects in toxicity tests for effects on growth or reproduction.

Effects on survival directly affect the hazard rate, which can be studied without detailed reference to energy budgets. It is perhaps better to call the model a hazard-based model, rather than a DEB-based one; the rich structure of the DEB model only comes into play for effects on growth and, especially, reproduction. The fact that changes in the hazard rate by toxicants beautifully link up with the effect of the aging process in the DEB model, can be considered as a happy coincidence that has little relevance for most standard toxicity tests. It becomes more relevant if the length of the toxicity test is not short with respect to the life span of the test organism. Evaluation of the consequences of effects on survival for population dynamics does involve the complete structure of the DEB model.

Effects on the population growth rate for dividing organisms such as algae only involve certain simple aspects of the DEB model, as is explained in Kooijman (1993). This is because the surface area/volume ratio only changes within a very restricted range during the cell cycle. This does not hold for animals such as water fleas and fish.

3 Examples of application of the DEB-based model

Examples of the application of the dynamic approach for toxicity tests for effects on survival, body growth, reproduction and population growth are given in Figures 1, 2, 3 and 4. All figures are composed from output files of the software package DEBtox, as provided in Kooijman & Bedaux (1996). It fits the response surface to all available data simultaneously, so the different curves in the concentration and exposure time profiles are linked and are not fitted independently.

Since the use of profile ln likelihood functions for the identifications of confidence intervals (here for the NEC) is not standard (probably because of the substantial amount of calculations that is involved), some guidance might be appropriate. The idea is that a confidence level is first selected on the horizontal axis in the left panel; the threshold

value for the ln likelihood function is then read off on the vertical axis, and the graph in the right panel is (mentally) intersected at this level. The intersection points represent the boundaries of the appropriate confidence interval. More than one confidence interval might result, because the ln likelihood function for the NEC can deviate substantially from a simple parabola, which is the shape that it should have if the large-sample theory for parameter estimation would apply. This is why this way of obtaining confidence sets is much more reliable than making use of the Asymptotic Standard Deviation (ASD, given in the parameter tables). The profile ln likelihood functions for the NEC in the toxicity tests on fish body growth and algae population growth are close to the expected parabola in these examples. DEBtox can also produce the numerical values of the interval estimates. Note that, generally, the NEC cannot easily be guessed from the concentration profiles, because this threshold corresponds with the EC0.t, while we have that NEC = EC0.∞.

The confidence intervals for the NEC in Figure 3 are not small. The main reason is in the uncertainty about the value of the elimination rate. If other information about this rate could be supplied (for instance from direct measurements, comparison with other toxicity tests or Quantitative Structure Activity Relationships), the confidence interval for the NEC can be reduced substantially.

DEBtox can also be used to test parameter values statistically and to extract all kinds of information about the effect surface, such as ECx values. This is not illustrated here.

4 Dynamic versus static approaches

4.1 Toxicity comparisons

Apart form comparing different models for the same data, we can and should compare data from different toxicity tests (different species of test organisms, different test chemicals, different endpoints). These comparisons yield some arguments that play a role in the comparison of the different models to the same data, and are, therefore, presented first.

4.1.1 Solubility in fat

The solubility in fat is a physical chemical property that is very relevant to the toxicity of a chemical. The linear effect model (i.e. the effect component of the DEB-based model), assumes that effective molecules operate independently. Since the ultimate number of molecules in an organism is directly proportional to the octanol/water partition coefficient, P_{ow}, the tolerance concentration and the NEC should be proportional to P_{ow}^{-1}, and the killing rate to P_{ow}. The latter relationship directly follows from the argument that the inverse of the killing rate is proportional to a tolerance concentration, as is obvious from the dimensions of the killing rate.

The symmetry argument states that the uptake flux depends on P_{ow} in the same way as the elimination flux depends on P_{wo}, while $P_{wo} = P_{ow}^{-1}$. This argument directly results in the expectation that the uptake rate is proportional to $\sqrt{P_{ow}}$ and the elimination rate is proportional to $1/\sqrt{P_{ow}}$. The logic is easily seen when we realize that the BCF equals the ratio of the uptake and the elimination rate, while it is proportional with P_{ow}.

Survival, Hazard model			ASD	Correlation coefficients		
Blank mortality rate	2.296e-011	d^{-1}	0.000			
No-effect concentration	190.2	$ug\ l^{-1}$	0.783	0.000		
Killing rate	0.009304	$l\ ug^{-1}\ d^{-1}$	0.001	0.000	0.081	
Elimination rate	7.019	d^{-1}	2.510	-0.000	0.134	-0.700
Deviance	31.41					

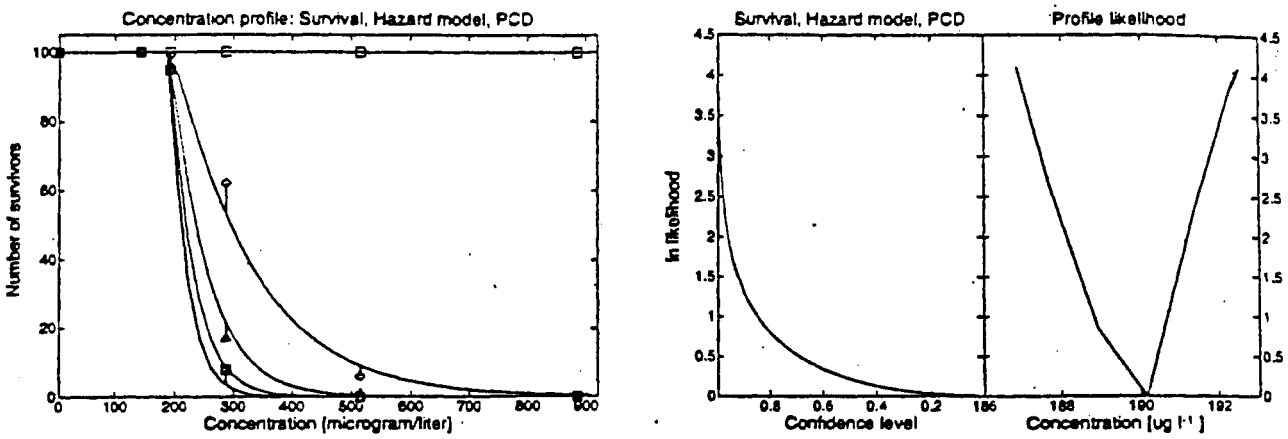


Figure 1: Effects of PCP on the survival of the fathead minnow *Pimephales promelas*. The profile ln likelihood function for the NEC, given in the graph on the left, has a non-typical shape, because DEBtox used few evaluation points. The reason is that the confidence interval of the NEC is here very small with respect to its value.

Body growth, Maintenance model			ASD	Correlation coefficients		
No effect concentration	111.3	$ug\ l^{-1}$	30.988			
Blank ultimate length	3.065	$g^{1/3}$	0.026	-0.298		
Tolerance concentration	546.8	$ug\ l^{-1}$	188.734	0.403	0.028	
Elimination rate	0.1059	d^{-1}	0.092	0.602	0.014	0.956
Initial length	1.56	$g^{1/3}$				
Von Bertalanffy growth rate	0.01	d^{-1}				
Energy investment ratio	1					
Mean deviation	0.01425	$g^{1/3}$				

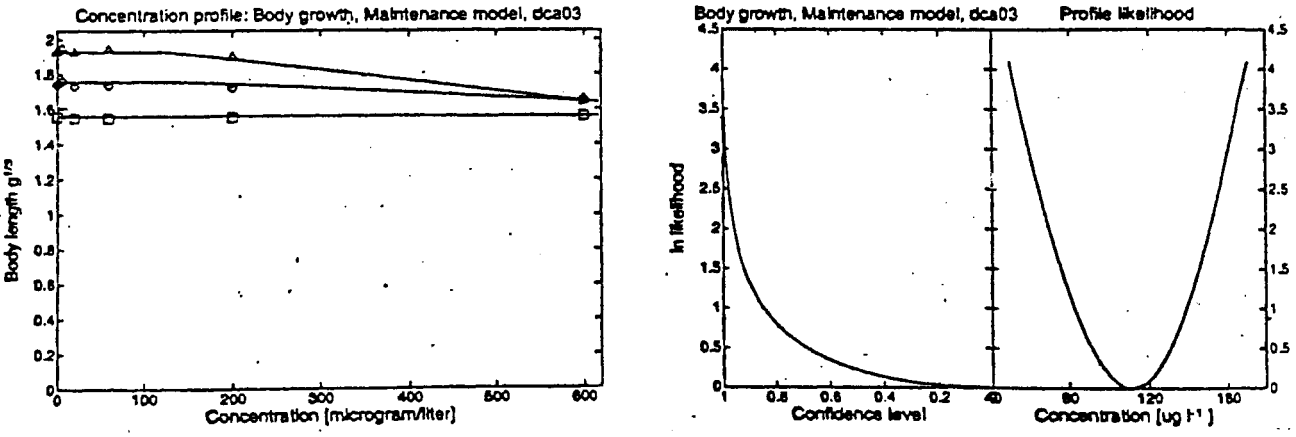


Figure 2: Effects of DCA on the body growth of the rainbow trout *Oncorhynchus mykiss*. Data from the OECD ring test 1988/9. The costs for maintenance has been selected as target parameter. The profile ln likelihood function for the NEC, given in the graph on the left, is close to its large sample shape: the parabola.

Reproduction, Growth model			ASD	Correlation coefficients		
No effect concentration	2.157	mg l^{-1}	0.117			
Tolerance concentration	0.702	mg l^{-1}	0.133	-0.193		
Maximal reproduction rate	30.94	No d^{-1}	0.359	-0.189	-0.031	
Elimination rate	0.6689	d^{-1}	0.124	0.122	0.950	-0.086
Von Bertalanffy growth rate	0.1	d^{-1}				
Scaled length at birth	0.13					
Scaled length at puberty	0.42					
Energy investment ratio	1					
Mean deviation	10.95					

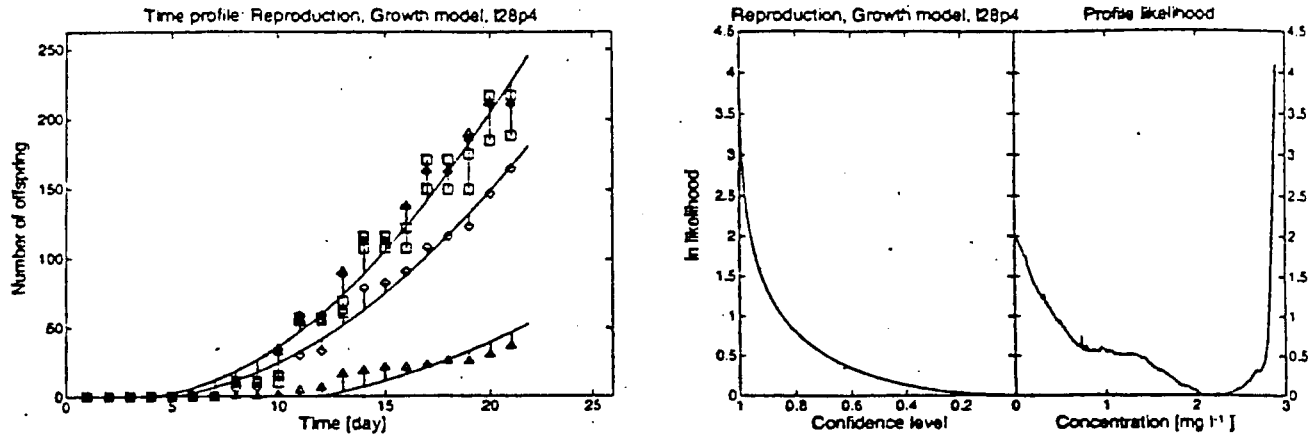


Figure 3: Effects of phenol on the reproduction of the water flea *Daphnia magna*. Data from the OECD ring test 1994/5. The costs for growth has been selected as target parameter. The little "teeth" in the profile ln likelihood function for the NEC, given in the graph on the left, are artifacts that resulted from numerical integrations to obtain the reproductive output. They are not typical; their occurrence depends on parameter values.

Population growth, Growth model			ASD	Correlation coefficients		
Inoculum size	5.663	•10^3 cells ml^{-1}	0.685			
Population growth rate	2.1	d^{-1}	0.063	-0.997		
No-effect concentration	0.7769	mg l^{-1}	0.031	0.118	-0.139	
Tolerance concentration	3.458	mg l^{-1}	0.307	-0.582	0.578	-0.524
Mean deviation	7.213	•10^3 cells ml^{-1}				

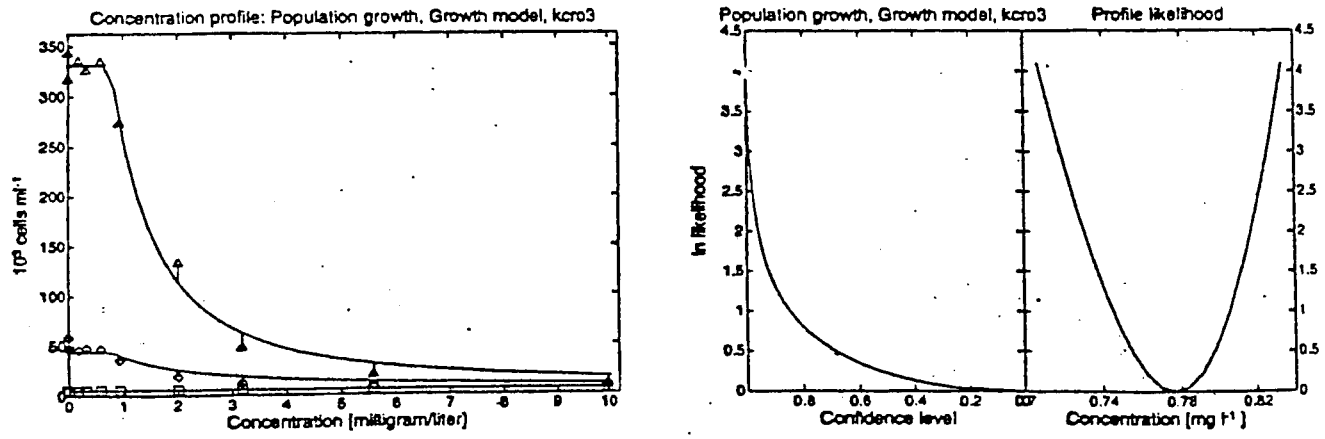


Figure 4: Effects of $K_2Cr_2O_7$ on the population growth rate of *Skeletonema costatum*. Data from the ISO ring test. The costs for growth has been selected as target parameter. The profile ln likelihood function for the NEC, given in the graph on the left, is close to its large sample shape: the parabola.

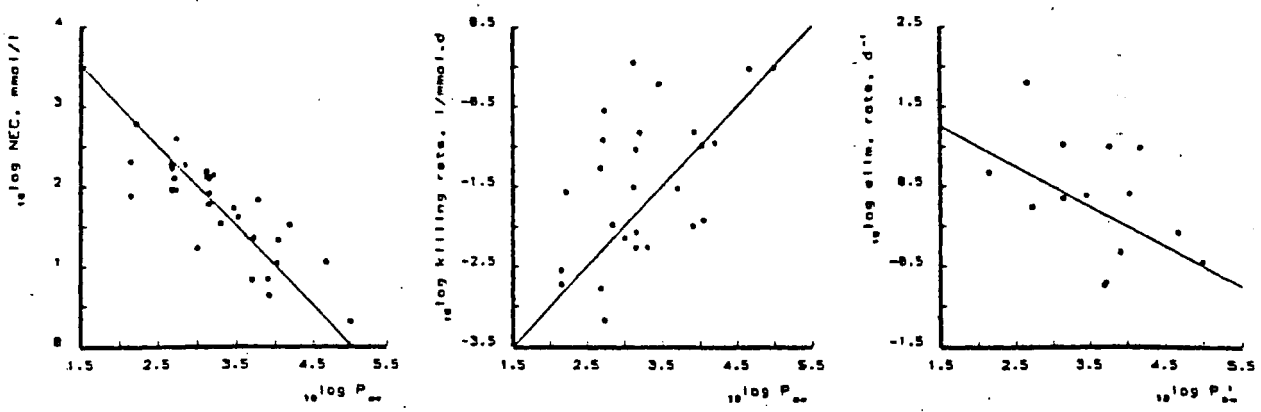


Figure 5: The NEC, killing rate and elimination rate of alkyl benzenes as a function of the octanol/water partition coefficient. The nitro benzenes have been excluded. The slopes of the lines, i.e. -1, 1 and -0.5, respectively, follow from simple theoretical considerations. The data are from the 4d toxicity tests on survival of the fathead minnow, as presented in Geiger et al 1985-1990. The partition coefficients were obtained from Richardson & Gangolli or calculated according to Rekker 1977.

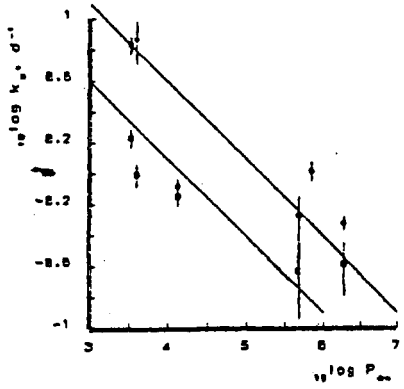


Figure 6: The elimination rate of alkyl benzenes as a function of the octanol/water partition coefficient for juvenile *Daphnia pulex* with a body length of 1 mm (•) and for adult ones of 3 mm (○). The lines correspond with the elimination rate $= 392/\sqrt{P_{ow}}$ d^{-1} for juveniles and $124/\sqrt{P_{ow}}$ d^{-1} for adults. The ratio between these elimination rates, i.e. 3.17, corresponds well with the ratio of the body lengths, as expected from simple theoretical considerations. The data are from the 2d toxicity tests on survival, as presented in Hawker & Connell 1985.

The strength of these simple relationships becomes obvious if we have the results of a toxicity test with a chemical with a P_{ow} of P_1 available and wonder about the toxicity of another test chemical with a P_{ow} of P_2. The relationships tell us to multiply the elimination rate with $\sqrt{P_1/P_2}$, and the NEC and tolerance concentration with P_1/P_2. These three statistics define the complete ECx.time behaviour of the second chemical. These expectations can help a lot in chosing the concentrations that are to be used in a toxicity test for the second chemical. This increase in efficiency helps to reduce the financial costs of a toxicity test.

Figure 5 presents the NEC, killing rate and elimination rate as functions of the octanol/water partition coefficients for alkyl benzenes. The theoretical predictions seem to apply, but the scatter, particularly for the elimination rate, is substantial.

The simplicity of the relationships of the DEB-based model parameters with the P_{ow} also helps to detect patterns in toxicity, comparing many test chemicals.

The EC50 of the logit model depends on the P_{ow} in a much more complex way, due to the fact that it depends on the standardized exposure time. Chemicals with a sufficiently

small P_{ow} reveal their toxic properties fully during the toxicity test, but chemicals with a large P_{ow} do not and the apparent toxicity will increase if the test would be extended. This can be understood on the basis of the relationship between the elimination rate and the P_{ow}, as mentioned above. The same problem applies to the NOEC. The allometric model EC50= $a(P_{ow})^b$ is usually fitted, but it is only based on empirical arguments.

4.1.2 Size of the organism

The EC50 and NOEC not only relate in a more complex way to the solubility in fat, compared to the dynamic parameters, but also to the body size of the test organism, by exactly the same argument. The larger the test organism, the longer it takes for effects to build up. This makes that the LC50s for daphnia and fish are difficult to compare. The comparison of the parameters of the DEB-based model is easy, because the NEC and the killing rate (or tolerance concentration) do not depend on body size. The elimination rate is inversely proportional to a volumetric length, on the assumption that the exchange rate between organism and environment are proportional to the surface area of the organism. It is the only parameter that depends on body size. Figure 6 illustrates that these simple considerations make sense.

4.2 Comparisons on the basis of parameters

The logit model has three parameters: the blank response, the EC50 (LC50) and the slope. The NOEC is inconsistent with the logit model. Nonetheless it is frequently presented with the logit parameters in practice. It can be viewed as a fourth parameter which is 'estimated' in a rather odd way. (It can take a very limited number of values and its estimation procedure has regrettable properties.) An extrapolation factor is used (except in population growth tests) to extrapolate to ultimate effects. This factor counts as a fixed parameter. The problem with this 'parameter' is that it relates to toxic effects, rather than to responses in the blank, which implies that it should not be the same for all chemicals.

The DEB-based model has four parameters; a blank response, a NEC, a tolerance concentration (or killing rate) and an elimination rate. In addition to this, for growth and reproduction tests only, it has three or four, respectively, parameters that are not estimated from the data. All these these fixed parameters refer to details in the description of what happens in the blank. The blank response in the test does not provide the proper information for the fixed parameters. These parameters are

- scaled length at birth, i.e. the length at birth as a fraction of the maximum length in the blank, if the test would continue in time. This parameter applies to the toxicity test for reproduction. That for body growth uses a related parameter: the actual length at the start of the experiment. Although this fixed parameter can be treated as a free parameter, DEBtox treats it as a fixed one, because the size at the start of the experiment is frequently not measured in the case of early life stage tests.
- scaled length at puberty, i.e. the length at puberty as a fraction of the maximum length in the blank, if the test would continue in time. This parameter only applies to

reproduction. Puberty is defined as the moment at which allocation to reproduction starts, which is somewhat earlier than the moment of first reproduction.

- energy investment ratio is a dimensionless parameter with a rather complex interpretation, which is described in the separate document. Realistic values for animals such as daphnia and fish are around 1. Numerical studies indicate that variations around this value have very little effect on the toxicity parameters.
- von Bertalanffy growth rate, with dimension time^{-1}. The value of this parameter is more important than of the previous one, but it can be measured easily and is known for hundreds of species.

The fixed parameters depend on the species, or even the strain, the culture conditions and details of the experimental protocol. If a particular laboratory has standardized its experimental protocol, these parameters need to be tuned only once, and not for each test. Part of the differences of toxicity results among laboratoria can be attributed to differences in fixed parameters. Good estimates can be given for the standard choices of species on the basis of existing ecophysiological data (Kooijman and Bedaux 1996).

The elimination rate in the DEB-based model stands for the rate at which effects build up during exposure. Its conceptual role is more or less comparable with the extrapolation factor in the static approach, but differs in a statistical sense. It is an ordinary model parameter, for which point-estimates as well as interval-estimates are available.

The slope parameter and the EC50 in the logistic model together play the same role as the tolerance concentration in the DEB-based model.

We can conclude that the DEB-based model has less parameters that are to be estimated from the results of the toxicity test than the logit model plus NOEC. On the basis of this measure for the complexity of a model, the simplest dynamic model is, therefore, simpler than the static one for all four toxicity tests.

4.3 Numerical comparisons

In this section we present numerical comparisons between the DEB-based model and the logit model plus the NOEC. Since the logit model only uses the response at the end of the experiment (apart from the algal growth inhibition test), we restrict the comparisons to ECx values for that exposure time. We compare the NOEC with the DEB-based EC0 for that exposure time, and not with the NEC, because the NEC relates to long-term exposure. The aim of this section is to compare statistics that are familiar to users of the static approach, not to show the potential of the dynamic one.

All calculations for the DEB-based model have been done with the software package DEBtox, as provided in Kooijman & Bedaux (1996). All calculations are based on nominal concentrations, except for the survival analyses, which are based on measured concentrations.

We fitted models for different modes of action of the compounds (i.e. target parameters) and noted that the differences in goodness of fit were generally small, and resulted in very similar values for the NEC.

Effects on survival

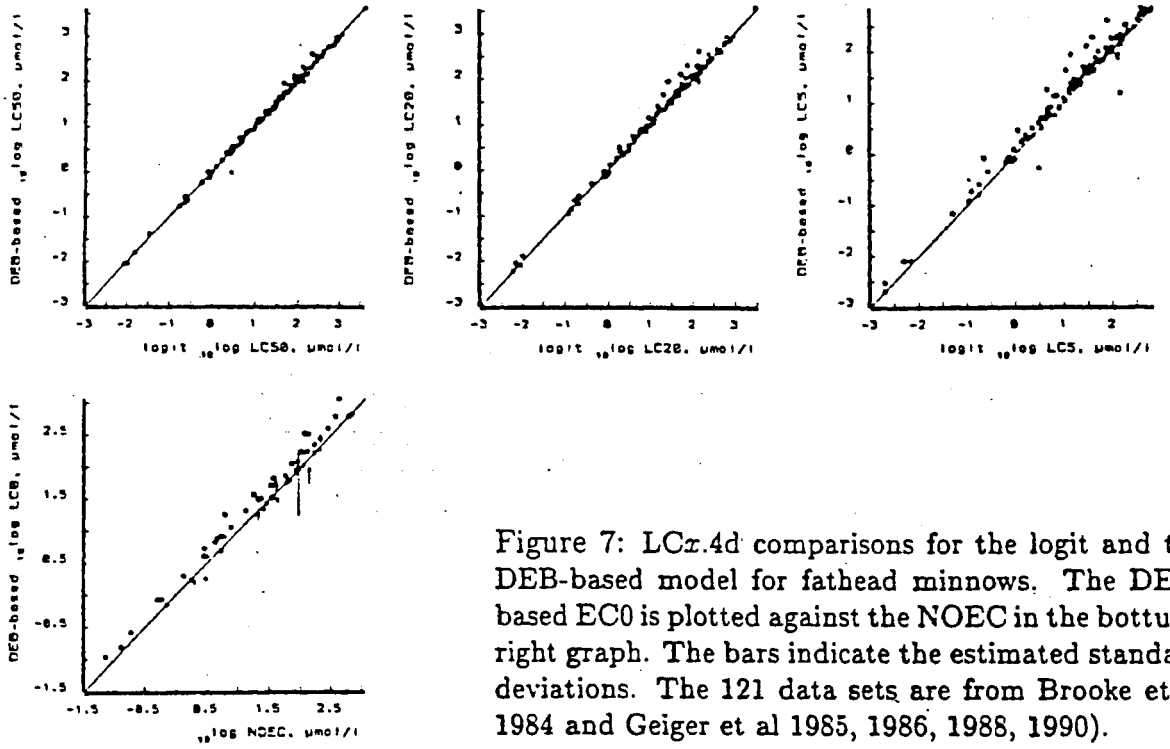


Figure 7: LCx.4d comparisons for the logit and the DEB-based model for fathead minnows. The DEB-based EC0 is plotted against the NOEC in the bottum-right graph. The bars indicate the estimated standard deviations. The 121 data sets are from Brooke et al 1984 and Geiger et al 1985, 1986, 1988, 1990).

4.3.1 Survival experiments

Figure 7 compares the LCx.4d values for fathead minnows on the basis of the logit and the DEB-based model for many different chemicals. The logit parameters have been estimated according to the maximum likelihood method, as described in Kooijman 1981. We can conclude that the LC50 and LC20 values are very similar for both models, but the LC5 values for the DEB-based model tends to be higher than that of the logit model. This is to be expected, because LC$x \to 0$ for $x \to 0$ for the logit model, but LC0.4d $\geq$ NEC for the DEB-based model. The fact that the LCx values of the logit and DEB-based models correlate well is not surprizing, because both models are fitted to the same data; very toxic chemicals result in low LCx values with both methods.

Effects on body growth

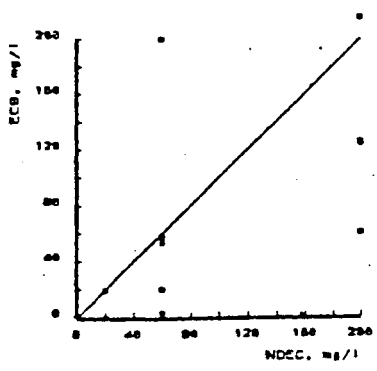


Figure 8: The EC0.28d for the DEB-based model for effects of DCA on the body growth of the rainbow trout is plotted against the NOEC on the basis of the Williams test for 11 toxicity tests of the OECD ringtest 1988. Effects on growth via effects on maintenance costs turned out to be the best fitting DEB-based model.

4.3.2 Body growth experiments

We used the data from the final ring test, organized by the OECD in 1988, with 3,4-dichloroaniline (DCA), for the rainbow trout *Oncorhynchus mykiss*. Since the volumetric lengths (i.e. the cubic root of weights) at the end of the toxicity tests differed little from those at the start, an analysis in terms of ECx values for the logit model is less appropriate: The EC50 would far exceed the highest tested concentration and the results would be most unreliable. We only present the results of the comparison of the NOEC with the EC0.28d in Figure 8.

The NOECs were identified on the basis of the test by Williams (Williams 1972), using a significance level of 5%. There were 16 individuals for each concentration of DCA.

The DEB-based model for effects on maintenance fitted best, although the differences in goodness of fit with the models for effects on growth and assimilation were usually small. The fixed parameters have been set at: initial length = mean initial length, energy investment ratio = 1, von Bertalanffy growth rate = 0.01 d^{-1}.

4.3.3 Reproduction experiments

We used the data from the final ring test, organized by the OECD in 1994/5, with 3,4-dichloroaniline (DCA), cadmium chloride and phenol. See Figures 9 and 10.

The NOEC values for these data were taken from Anonymous (1995), as listed for the case that the reproduction of females that die during the test are excluded.

The logit model has been fitted with SYSTAT version 5.02 using non-linear regression on the number of juveniles per adult for individuals that did not die in 21 days. The ECx values, and their standard deviations have been obtained via reparametrization of the logit model.

The data from DEBtox represent the cumulated total reproduction per living female, including all observation times. The reproduction of a female that died during the test has been included. No delay of reproduction could be observed for cadmium, and we selected the model for effects on the hazard of the ovum as the best fitting model for direct effects on reproduction. This model is rather similar to the other direct effect model, i.e. effect on costs per young. Where difference in goodness of fit for both models was relatively large (but still small absolutely), the hazard model fitted best. Some delay of reproduction could be observed for DCA and phenol, and we selected the model for effects on the growth costs as the best fitting model for indirect effects on reproduction. Since no data on the size of the adults are available, we could not test the predicted effects on growth. The fixed parameters have been set to the default settings of DEBtox: i.e. scaled length at birth = 0.13, scaled length at puberty = 0.42, energy investment ratio = 1, von Bertalanffy growth rate = 0.1 d^{-1}.

The growth model (or hazard model respectively) could be fitted to all data sets without problems, using DEBtox. The NECs and tolerance concentrations for data sets without effect of the toxicant were large with huge standard errors, as could be expected. We did not include data sets into the comparisons where the EC50.21d was much larger than the highest tested concentration.

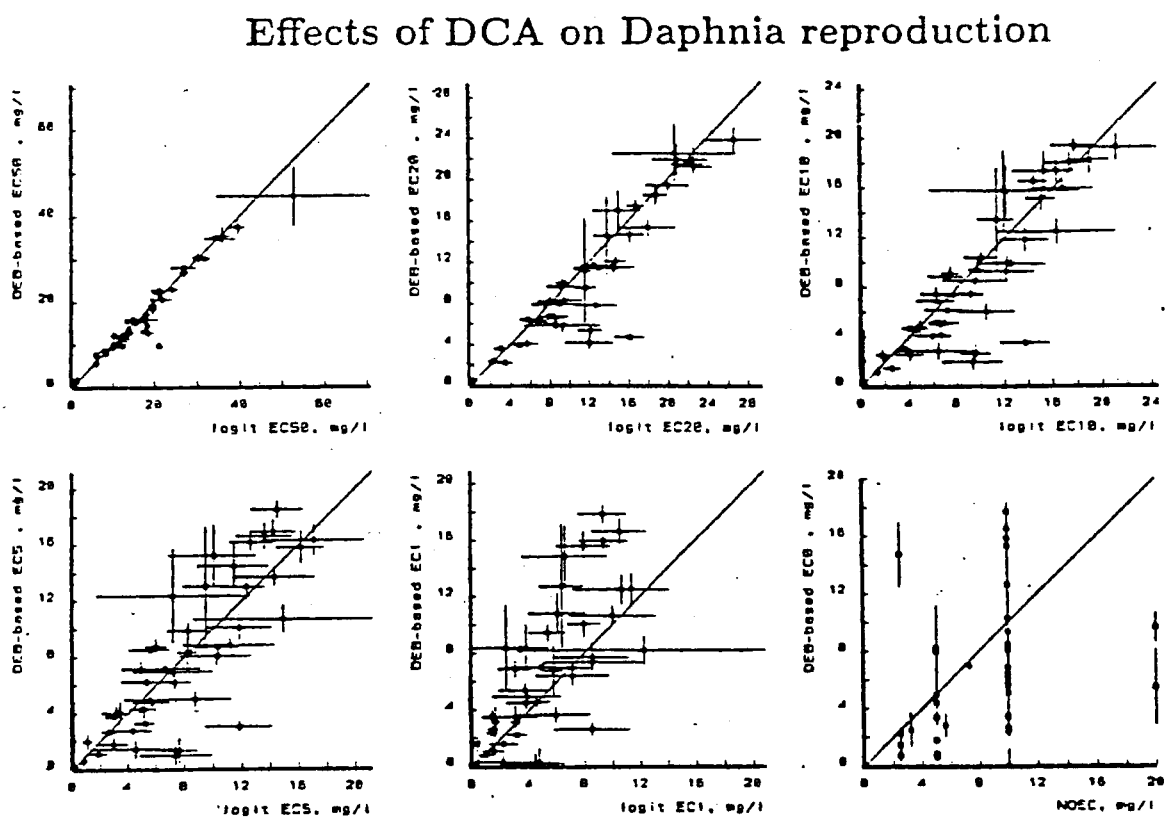


Figure 9: ECx.21d comparisons for the logit and the DEB-based model for the daphnia reproduction test on DCA. The DEB-based EC0.21d is plotted against the NOEC. The bars indicate the estimated standard deviations. The data sets are from the final ring test of the OECD 1994/5. Four NOECs could not be obtained, and have been set to zero in the graph.

The Figure 9 and 10 present the comparisons of the ECx values and the NOECs for the chemicals DCA, cadmium and phenol. The ECx values are quite comparable for both models if $x \geq 5$. The standard deviations of the ECx values of the logit model tend to be somewhat larger than for the DEB-based model. The EC1 values of the logit model tend to be lower than that of the DEB-based one. This can be expected, because the EC$x \rightarrow 0$ if $x \rightarrow 0$ for the logit model, while EC0≥NEC for the DEB-based model. We also see that the scatter in the NOECs is larger than the scatter in the EC0s for two of the three chemicals.

4.3.4 Population growth experiments

We analysed the data on the effects of 3,5 dichloro-phenol (DCP) and potassium dichromate from the ISO ring test with the diatoms *Skeletonema costatum* and *Phaeodactylum tricornutum* (marine algal growth inhibition test, ISO 10253, Geneva 1995).

The logit model was fitted to the data using non-linear regression, as described in Kooijman et al (1983), where the population growth rate decreases logistically with the concentration of test chemical. The standard deviations of the ECx values were obtained from those of the EC50, the slope and the covariance of both parameters, on the basis of

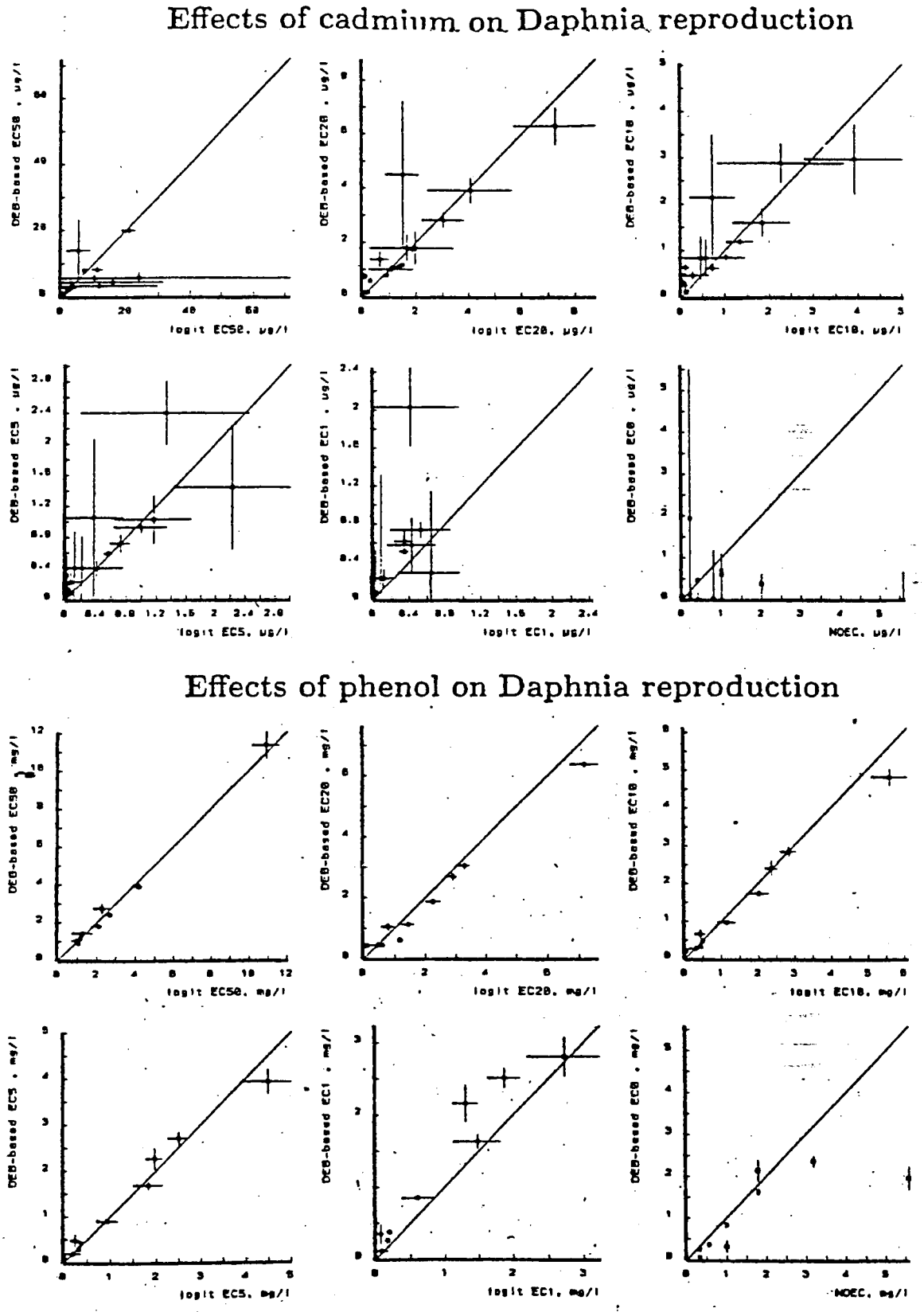


Figure 10: ECx.21d comparisons for the logit and the DEB-based model for the daphnia reproduction test on cadmium and phenol. The DEB-based EC0.21d is plotted against the NOEC. The bars indicate the estimated standard deviations. The data sets are from the final ring test of the OECD 1994/5.

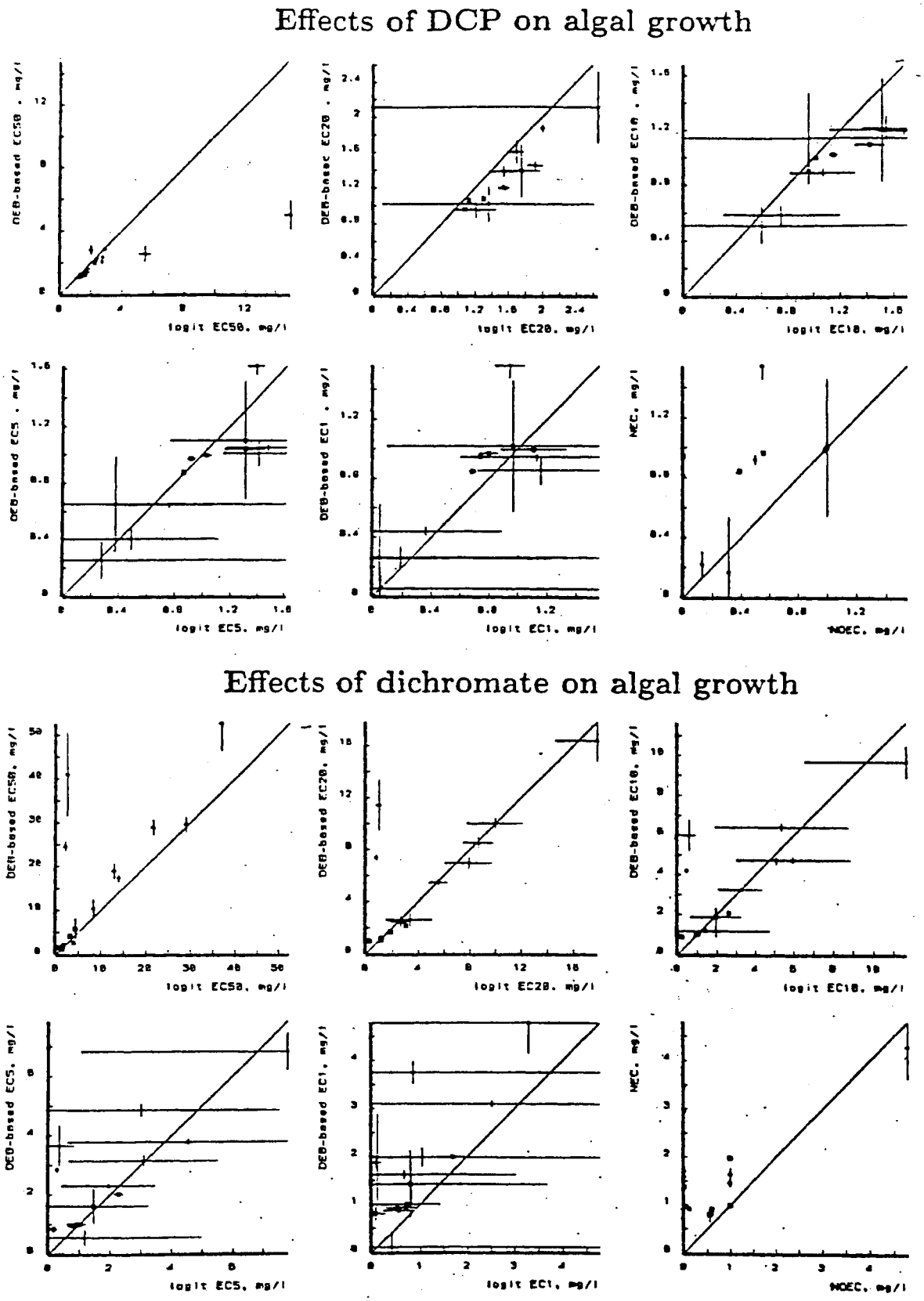


Figure 11: ECx comparisons for the logit and the DEB-based model for the alga growth inhibition test on dichromate and DCP. The DEB-based NEC is plotted against the NOEC in the bottum-right graph. The bars indicate the estimated standard deviations. The data sets are from ISO ring tests.

the formulae presented in Kooijman and Bedaux (1996).

The DEB-based model for effects on growth costs fitted best for dichromate, while that for effects on the hazard rate fitted best for DCP. Since the DEB-based model assumes that toxicant kinetics is fast with respect to population growth, we have that EC0 = NEC.

Figures 11 compare the ECx values and the NOECs. The two data sets for dichromate, where the EC50 and the EC20 are small for the logit model and large for the DEB-based one, have already been indicated by Hanstveit (1991) as outliers. For DCP we have the opposite situation: two data sets for which the EC50 is large for the logit model and small for the DEB-based one. We see that the logit model has relatively small standard deviations for the EC50, but they rapidly become larger for the smaller effect levels to the extent that they become meaningless in quite a few data sets.

No proper NOEC could be identified in 4 of the 12 data sets for DCP and 6 of the 15 data sets for dichromate. These values have been set to zero in the graphs. The NEC was estimated to be zero in 1 data set for DCP and 1 for dichromate.

4.4 Advantages of the dynamic approach

- The DEB-based model allows the estimation of the NEC, which has good statistical properties, including an interval estimate.
- The dynamic approach needs no extrapolation factor to arrive at long-term effects. The elimination rate plays this role.
- The comparison of results of different toxicity tests is easy on the basis of the DEB-based model, because (i) its parameters are independent of the exposure time, (ii) it gives simpler QSARs and (iii) simpler relationships with the body size of the test animals.
- The static parameters (ECx values) can be obtained from dynamic ones, but not vice versa, which illustrates that the dynamic approach extracts more information from the same toxicity data.
- The dynamic approach uses all available data, while the static approach uses only the data at the end of the exposure. It has less parameters that are to be estimated from the data than the static approach.
- Additional information, such as the elimination rate, derived from toxico-kinetic data, can readily be used in the dynamic approach.
- The effects of repeated exposures and of pulse exposures can be evaluated readily, because of the dynamic properties of the DEB-based model.
- The toxicity of mixtures of compounds can readily be evaluated, including interactions between different compounds, due to the linearity of effect component of the DEB-based model.
- The effects of mutagenic compounds and ionogenic radiation can be quantified via effects on the target parameter 'ageing acceleration' (see separate document).

- Environmental risk assessment concerns effects of toxicants on "natural" populations, not on individuals in a laboratory. To evaluate the consequences of toxic effects on individuals for population dynamics, we have to know how survival and reproduction change during the lifetime of an organism. This requires a dynamic approach.
- The DEB-based model is mechanistic, based on biological ideas. It allows a series of models from simple (routine testing) to complex (research) by changing the kinetics component.

4.5 Disadvantages of the dynamic approach

- The dynamic approach requires rather advanced numerical techniques for statistical evaluations. This problem is solved by the software package DEBtox.
- The blank component in the dynamic approach is more complex than in the static approach and involves biological knowledge. This knowledge also must be used in the application of the toxicity results in environmental risk assessment. Fixed parameters for the blank response reflect this biological realism. Supplementary eco-physiological data are required for each species of test organism to determine the values. These data are available for the species that are frequently used, and appropriate values for the fixed parameters are known.
- The different modes of action of the various compounds relate to different target parameters, which hampers the comparison of the toxicity of such compounds to some extent. The proper identification of the mode of action is not always feasible with the present experimental design. Additional observations, such as body length at the end of the daphnia reproduction experiment and/or the measurement of the feeding rate, would help substantially. Numerical results indicate that the proper identification of the mode of action is not essential for a reliable estimation of the NEC.
- The dynamic approach does make assumptions about the effects, assumptions that can be wrong. One should realize, however, that every approach is based on assumptions. The use of a fixed extrapolation factor to transform the estimated toxicity into a long-term toxicity in the static approach, for instance, assumes that this factor is the same for all compounds. The assumptions are more explicit and more realistic in the dynamic approach.

References

Anonymous 1995. Draft report of the final ring test of the *Daphnia magna* reproduction study. Draft OECD test guideline 202.

Brooke, L.T., Geiger, D.L., Call, D.J. & Northcott, C.E. 1984. *Acute toxicities of organic chemicals to fathead minnow (*Pimephales promelas*)* Vol 1. Center for Lake Superior Environmental Studies. Univerity of Wisconsin-Superior, USA.

Geiger, D.L., Brooke, L.T. & Call, D.J. 1990. *Acute toxicities of organic chemicals to fathead minnow*

(Pimephales promelas) Vol 5. Center for Lake Superior Environmental Studies. Univerity of Wisconsin-Superior, USA.

Geiger, D.L., Call, D.J. & Brooke, L.T. 1988. *Acute toxicities of organic chemicals to fathead minnow (*Pimephales promelas*)* Vol 5. Center for Lake Superior Environmental Studies. Univerity of Wisconsin-Superior, USA.

Geiger, D.L., Northcott, C.E., Call, D.J. & Brooke, L.T. 1985. *Acute toxicities of organic chemicals to fathead minnow (*Pimephales promelas*)* Vol 2. Center for Lake Superior Environmental Studies. Univerity of Wisconsin-Superior, USA.

Geiger, D.L., Poirier, S.H., Brooke, L.T. & Call, D.J. 1986. *Acute toxicities of organic chemicals to fathead minnow (*Pimephales promelas*)* Vol 3. Center for Lake Superior Environmental Studies. Univerity of Wisconsin-Superior, USA.

Hanstveit, A.O. 1991. The results of an international ring test of marine algal growth inhibition test according to ISO/DP 10253. TNO Report 91/236.

Hawker, D.W. & Connell, D.W. 1986. Bioconcentration of lipophilic compounds by some aquatic organisms. *Ecotox. Environ. Safety* **11**: 184-197.

Kooijman, S.A.L.M. 1981. Parametric analyses of mortality rates in bioassays. *Water Res.* **15**: 107-119.

Kooijman, S.A.L.M. & Bedaux, J.J.M. 1996. *The analysis of aquatic toxicity data.* VU University Press, Amsterdam (pp 160 + floppy)

Kooijman, S.A.L.M., Hanstveit, A.O. & Oldersma, H. 1983. Parametric analyses of population growth in bioassays. *Water Res.* **17**: 727-738.

Rekker, R.F. 1977. The hydrophobic fragmental constant. Its derivation and application. A means of characterizing membrane systems. In: Pharmacochemistry Library. Nauta, W.Th. & Rekker, R.F. (eds), Elsevier, Amsterdam.

Richardson, M.L. & Gangolli, S. 1995. *The dictionary of substances and their effects.* Publication of the Royal Society of Chemistry, Cambridge, CB4 4WF

Williams, D.A. 1972. The comparison of several dose levels with a zero dose control. *Biometrics* **28**: 519-531.

Appendix: Publications on DEB-based effect models

Individual-level

Bedaux, J.J.M. & Kooijman, S.A.L.M. 1994. Statistical analysis of bioassays, based on hazard modeling. *Envir. & Ecol. Stat.* **1**: 303-314.

Haren, R.J.F.van, Schepers,H.E. & Kooijman, S.A.L.M. 1994. Dynamic Energy Budgets affect kinetics of xenobiotics in the marine mussel *Mytilus edulis*. *Chemosphere* **29**: 163-189

Kooijman, S.A.L.M. 1996. An alternative for NOEC exists, but the standard model has to be replaced first. *Oikos* **75**: 310-316.

Kooijman, S.A.L.M. 1996. Process-oriented descriptions of toxic effects. In: Schüürmann, G. and Markert, B. (eds) *Ecotoxicology*, John Wiley. (to appear)

Kooijman, S.A.L.M. & Bedaux, J.J.M. 1996. Some statistical properties of estimates of no-effects concentrations. *Water Res.* **30**: 1111-1111.

Kooijman, S.A.L.M. & Bedaux, J.J.M. 1996. Analysis of toxicity tests on *Daphnia* survival and reproduction. *Water Res.* **30**: 1711-1723.

Kooijman, S.A.L.M. & Bedaux, J.J.M. 1996. Analysis of toxicity tests on fish growth. *Water Res.* **30**: 1633-1644.

Kooijman, S.A.L.M. & Bedaux, J.J.M. 1996. *The analysis of aquatic toxicity data*. UV University Press, Amsterdam (pp 160 + floppy)

Kooijman, S.A.L.M., Bedaux, J.J.M. & Slob, W. 1996. No-Effect Concentration as a basis for risk assessment. *Risk Analysis* **16** (4) (to appear)

Kooijman, S.A.L.M., Hanstveit, A.O. & Nyholm, N. 1996. No-effect concentrations in alga growth inhibition tests. *Water Res.* **30**: 1625-1632.

Kooijman, S.A.L.M. & Haren, R.J.F.van 1990. Animal energy budgets affect the kinetics of xenobiotics. *Chemosphere* **21**: 681-693.

Population level

Hallam, T.G., Lassiter, R.R. & Kooijman, S.A.L.M. 1989. Effects of toxicants on aquatic populations. In: Levin, S.A., Hallam, T.G. & Gross, L.F. (eds) *Mathematical Ecology*. Springer, London pp 352-382.

Kooijman, S.A.L.M. 1991. Oecotoxicologische risico-evaluatie. In: Straalen, N.M.van & Verkleij, J.A.C. (eds) *Leerboek Oecotoxicologie*: 348-356. VU University Press, Amsterdam.

Kooijman, S.A.L.M. 1991. Effects of feeding conditions on toxicity for the purpose of extrapolation. *Comp. Biochem. Physiol.* **100c(1/2)**: 305-310.

Kooijman, S.A.L.M. 1985. Toxicity at population level. In: Cairns, J. (ed). *Multispecies toxicity testing*: 143-164. Pergamon Press, N.Y.

Kooijman, S.A.L.M. 1988. Strategies in ecotoxicological research. *Environmental Aspects of Applied Biology* **17** (1): 11-17.

Kooijman, S.A.L.M., Hanstveit, A.O. & Hoeven, N.van der 1987. Research on the physiological basis of population dynamics in relation to ecotoxicology. *Wat. Sci. Tech.* **19**: 21-37.

Kooijman, S.A.L.M. & Metz, J.A.J. 1983. On the dynamics of chemically stressed populations; The deduction of population consequences from effects on individuals. *Ecotox. Envir. Safety* **8**: 254-274.

ANNEX 9

The Dynamic Energy Budget (DEB) model

S.A.L.M. Kooijman, 12 sept 96

This short note introduces the Dynamic Energy Budget (DEB) model which specifies the rules for uptake and use of food for ectothermic ("cold-blooded") animals. The term 'Dynamic' refers to the change of the energy budget during the life history of an animal, see Figure 1. Three stages are distinguished: the embryo (which does not eat), the juvenile (which eats, but does not reproduce) and the adult (which eats and reproduces). With minor modifications, the model also applies to endothermic ("warm-blooded") animals and unicellulars (including bacteria) that are limited in growth by a single resource.

The diagram in Figure 2 presents the fluxes of energy through an animal, as conceived in the DEB model. Energy is extracted from food and added to the reserves, i.e. a combination of carbohydrates, lipids and proteins. Energy in the reserves is used for four destinations, which can be combined into two groups of two: growth (i.e. increase in structural biomass, mainly in the form of proteins) plus somatic maintenance (including activity, protein turnover, etc) and maturation (i.e. development, the increase in the state of maturity) plus maturity maintenance (i.e. maintaining the acquired state of maturity). Adults do not longer invest into maturation, but into reproduction. The various destinations only compete within each group. So, the animal ceases growth when the energy allocated to growth plus somatic maintenance is fully required for somatic maintenance. Under these conditions, it can continue to reproduce (if it is an adult), because reproduction is not in the same group of destinations. Likewise, reproduction ceases when the energy allocation to reproduction plus maturity maintenance is fully required for maturity maintenance.

The rules, presented as axioms in Table 1, quantify the fluxes that are shown in Figure 2. Each of these axioms can be justified mechanistically, and has been tested against experimental data for a wide variety of species. These simple rules have a myriad of implications for suborganismal organization and population dynamics. For instance, the energy costs of an egg and its incubation time follow directly from these rules. Although the rules define energy fluxes, all mass fluxes, including respiration (i.e. oxygen use or carbon

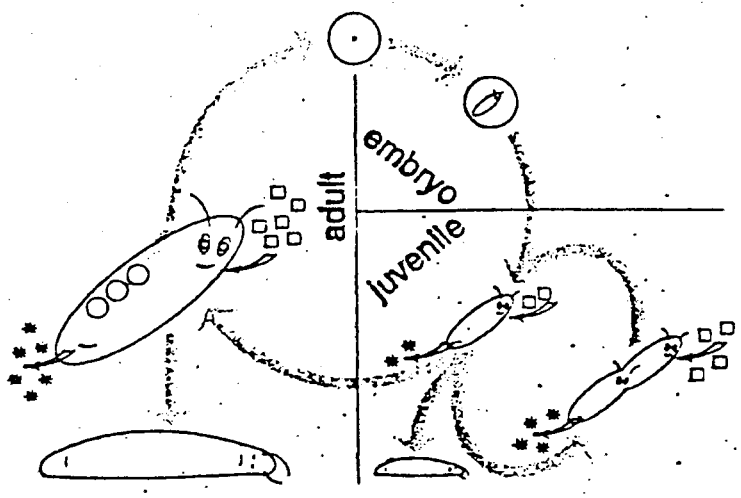


Figure 1: Dynamic Energy Budget theory quantifies the energetics as it changes during life history. The key processes are feeding, digestion, storage, maintenance, growth, development, reproduction and aging. Dividing organisms, such as microbes, are included by conceiving them as juveniles.

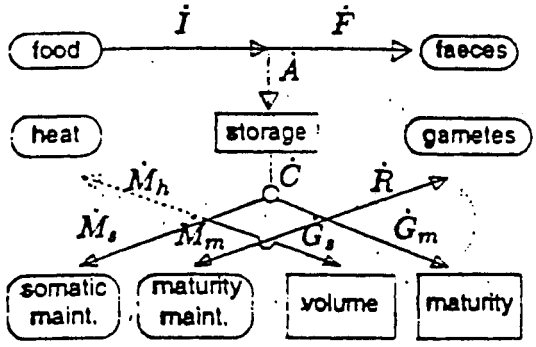


Figure 2: Energy fluxes through an animal: $\dot{I}$ ingestion (uptake), $\dot{F}$ defecation, $\dot{A}$ assimilation, $\dot{C}$ catabolic, $\dot{M}_s$ somatic maintenance, $\dot{M}_m$ maturity maintenance, $\dot{M}_h$ heating (endotherms), $\dot{G}_s$ somatic growth, $\dot{G}_m$ maturation, $\dot{R}$ reproduction. The rounded boxes indicate sources or sinks. All fluxes contribute a bit to dissipating heat, but this is not indicated in order to simplify the diagram.

dioxide production) and the rate of nitrogen waste (ammonia, urine) also follow from these rules, via the conservation law for mass. The rules give an explanation for the observed increase in respiration coupled to the feeding process (this previously poorly understood phenomenon is called the 'specific dynamic action'). It can be shown that the rules provide a theoretical basis for the widely applied method of indirect calorimetry, where measurements of oxygen use, carbon dioxide production and nitrogen waste are used to obtain the flux of dissipating heat.

The DEB model specifies the uptake and use of food by an animal as a dynamic system with three state variables (volume of structural biomass, amount of reserves and cumulated damage) and 11 parameters:

L_b	length at birth	L_p	length at puberty	K	saturation constant
$\{\dot{I}_m\}$	max. spec. ingestion rate	$\{\dot{A}_m\}$	max. spec. assim. rate	$\tilde{p}_a$	ageing acceleration
$[\dot{M}]$	spec. somatic maint. costs	$[G]$	spec. growth costs	$[E_m]$	max. spec. reserves
κ	$\frac{\text{somatic maint.+growth costs}}{\text{catabolic energy}}$	q	overhead costs of reprod.		

Although the number of parameters might seem large, it is in fact extremely small in view of the number of processes that are specified. Only a small selection of these parameters is involved in the description of any particular measured variable. If we evaluate the expression for size as a function of age, for instance, we know beforehand that parameters with energy in their dimensions will occur only as ratios, such that the dimension energy drops out. This is because energy is not involved directly in size measurements. We need to know the value of a parameter that has energy in its dimensions only if we want to describe energies.

Three compound parameters frequently appear in expressions for physiological quantities (cf Figure 3): The maintenance rate constant, $\dot{m} \equiv [\dot{M}]/[G]$ (dimension: time^{-1}); the energy conductance, $\dot{v} \equiv \{\dot{A}_m\}/[E_m]$ (dimension: length time^{-1}); the energy investment ratio, $g \equiv [G]/\kappa[E_m]$ (dimension: none).

Figure 3 presents the feeding-at-length, respiration-at-length, growth-at-age and reproduction-at-age of the waterflea *Daphnia magna* for the situation of constant food density and temperature. This species, like most other species of animal, hardly changes its shape during growth, which implies that its surface area is proportional to the squared volumetric length and its volume to the cubed length. The four relationships in Figure 3 cover the major processes of uptake and use of food. The expressions, which follow from the set of rules of Table 1, show how the (compound) parameters in the description of these

relationships depend on the feeding conditions. The scaled functional response f (defined as the ratio of the ingestion rate and the maximum one for an animal of that size) is under experimental control. Length-at-age turns out to follow the von Bertalanffy growth curve. By choosing different feeding levels, the von Bertalanffy growth rate (which is a compound parameter) and the ultimate length (another compound parameter) change in a particular way. This information can be used to estimate the compound parameters ($\dot{m}$, $\dot{v}$ and g) that are involved. These compound parameters can also be estimated from data such as the specific rate of weight decrease during starvation, respiration ontogeny during the embryonic period and survival probability-at-age.

The most far reaching and spectacular implications of the rules are the inter-specific body-size-scaling relationships. These relationships give trends in parameter values as they covary over different species (bacteria to whales). The 11 parameters can be classified in two groups. One group of parameters does not depend on body size, while the other group does depend on body size in a simple and predictable way: these parameters are proportional to the volumetric length, i.e. the cubic root of the body volume. Deviations of parameter values from these trends reflect ecophysiological adaptations of that species. All physiological variables that can be written as functions of the parameters can, for this reason, also be written as functions of (maximum) body size. Many of these functions have been worked out, tested against data and found to be realistic. Among them is the respiration rate, which turns out to be a weighted sum of squared and cubic (volumetric) length. This is very similar to empirical relationships, that indicate that respiration is approximately proportional to body mass to the power 0.75. The DEB model solves the long standing problem of understanding this empirical relationship.

The intra-specific body-size-scaling relationships (where we have just one set of parameters to describe the processes of food uptake and use) are fundamentally different from inter-specific body-size-scaling relationships (where we have 'sloppy' trends in parameter values among species). Hence, the fact that respiration, as it increases during the growth of an indivudual, turns out to be a weighted sum of squared and cubic length, just like for inter-specific comparisons, is merely coincidence. The volume-specific respiration decreases with body volume during life, because of the decreasing investment into growth.

A full description of the theory can be found in:

Kooijman, S.A.L.M. 1993. *Dynamic Energy Budgets in Biological Systems. Theory and applications in ecotoxicology.* Cambridge University Press, ISBN 0-521-45223-6, 350 pp.

Table 1: Key assumptions of the DEB model that specify a dynamical system with the state variables structural body mass, reserves and cumulated damage.

- If the investment into maturation exceeds a given threshold value, the organism changes its stage, i.e. it switches from the embryonic stage to the juvenile stage by initiating the feeding process, or from the juvenile stage to the adult stage by ceasing maturation and initiating the production of gametes (eggs, sperm).
- Food uptake is proportional to surface area and depends hyperbolically on food density (i.e. Holling type II functional response).
- The specific reserve dynamics is first order, with a rate that is inversely proportional to the volumetric length.
- The allocation to somatic maintenance plus growth (i.e. increase in structural biomass) is a fixed fraction of the energy drain from the reserves, which further includes maturity maintenance plus maturation or reproduction. This rule is called the κ-rule.
- Homeostasis of structural biomass and reserves, i.e. their chemical composition does not change, despite changes in the chemical composition of the environment. Since the amount of reserves can change relative to the structural biomass, certain changes in the chemical composition of the individual as a whole are possible. The homeostasis assumption implies that the following items are constant

  food-energy conversion, although it depends on the type of food

  volume-specific maintenance costs (both somatic and maturity)

  volume-specific growth costs
- The hazard rate is proportional to the accumulated damage

  the damage production is proportional to the changed DNA

  the DNA change is proportional to the use of oxygen
- The initial conditions are given by

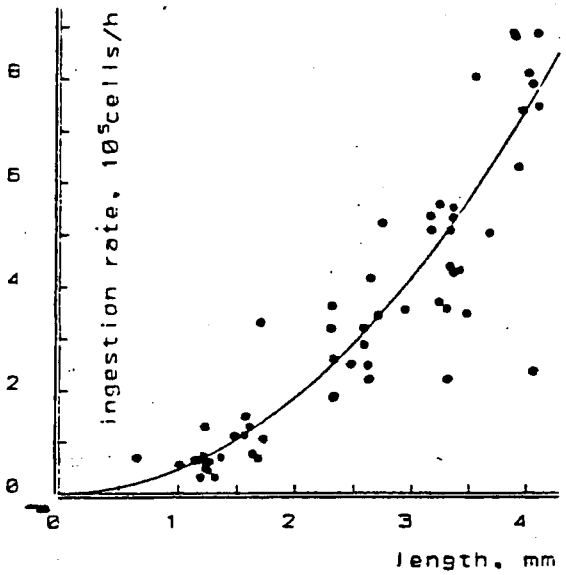
  The initial structural biomass is negligibly small.

  The reserve density at birth equals that of mother at egg laying.

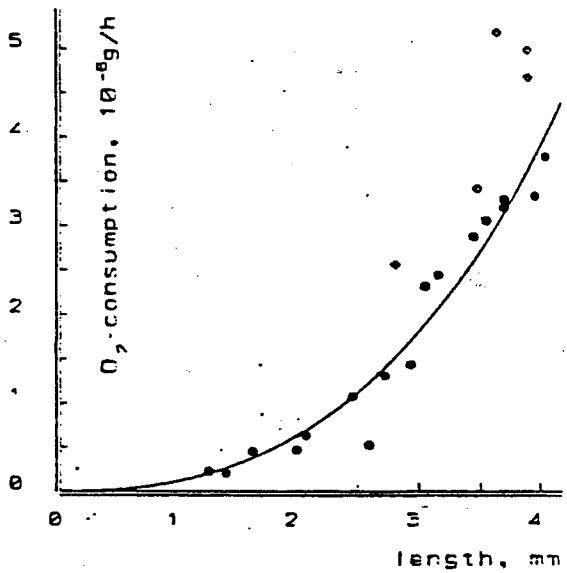
  The initial damage is negligibly small.

Figure 3: Investment into maturation and the κ-rule for allocation of energy from reserves solves the following puzzle: *Daphnia magna* starts to reproduce upon exceeding 2.5 mm body length. Reproduction takes about 80% of the budget. Where does this energy come from? Ingestion or respiration is not rapidly increased at this size, nor is growth reduced. The rules in Table 1 imply the expressions presented above the graphs, where L stands for body length, a for age, and the compound parameters are given in the text.

Ingestion $\propto fL^2$

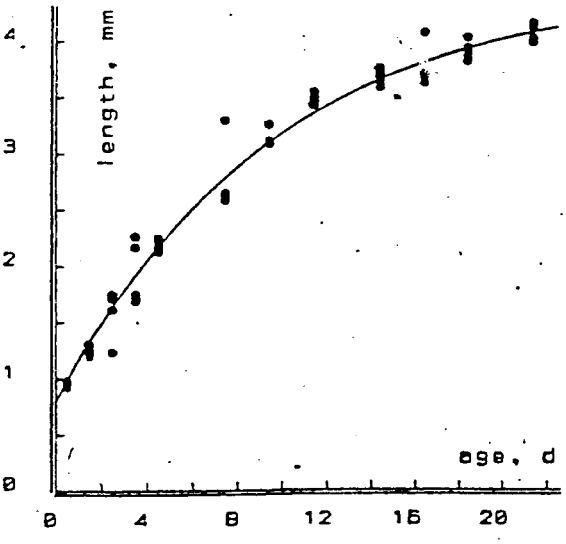


Respiration $\propto \dot{v}L^2 + \dot{m}L^3$

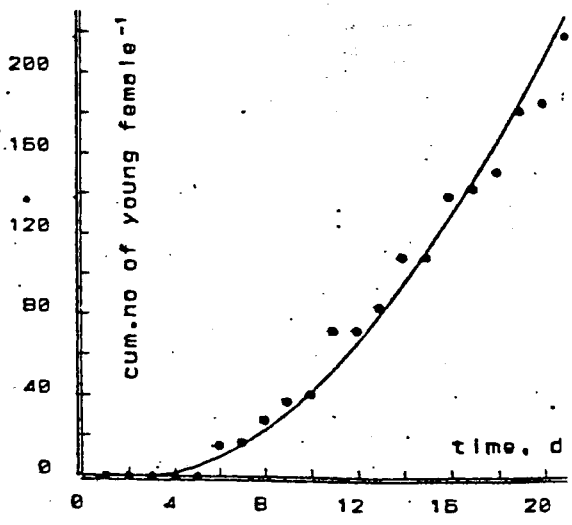


Growth: $\frac{d}{dt}L = \dot{\gamma}(L_\infty - L)$

$\dot{\gamma} = \frac{\dot{m}g}{3(f+g)}$; $\quad L_\infty = \frac{f\dot{v}}{g\dot{m}}$



Reproduction $\propto \dot{v}L^2 + \dot{m}L^3 - (1 + g/f)\dot{m}L_p^3$



Appendix: Publications on the DEB-model

General and individual-level

Evers, E.G. & Kooijman, S.A.L.M. 1989. Feeding and oxygen consumption in *Daphnia magna*; A study in energy budgets. *Neth.J.Zool.* 39: 56-78.

Haren, R.J.F. 1995. Application of Dynamic Energy Budgets to xenobiotic kinetics in Mytilus edulis and population dynamics of Globodera pallida. PhD-thesis, Vrije Universiteit, june 1995.

Haren, R.J.F. van & Kooijman, S.A.L.M. 1993. Application of the dynamic energy budget model to *Mytilus edulis* (L). *Neth. J. Sea Res.* 31: 119-133.

Konarzewski, M. & Kooijman, S.A.L.M. 1994. Models for growth in precocial and altricial birds. *J. Ornith.* 135 (3): 324

Konarzewski, M., Kooijman, S.A.L.M. & Ricklefs, R.E. 1996. Models for avian growth and development. In: Starck, J.M. & Ricklefs, R.E. (eds) *Avian growth and development.* Oxford University Press (to appear)

Kooijman, S.A.L.M. 1986. Energy budgets can explain body size relations. *J. Theor. Biol.* 121: 269-282.

Kooijman, S.A.L.M. 1986. What the hen can tell about her egg; Egg development on the basis of budgets. *J. Math. Biol.* 23: 163-185.

Kooijman, S.A.L.M. 1988. The von Bertalanffy growth rate as a function of physiological parameters; A comparative analysis. In: Hallam, T.G., Gross, L.J., & Levin, S.A. (eds) *Mathematical ecology*: 3-45. World Scientific, Singapore.

Kooijman, S.A.L.M. 1993. *Dynamic Energy Budgets in Biological Systems. Theory and applications in ecotoxicology.* Cambridge University Press, 350 pp.

Kooijman, S.A.L.M. 1994. Effects of temperature on birds. In: Hagemeijer, E.J.M. & Verstrael, T.J. *Bird Numbers 1992. Distribution, monitoring and ecological aspects.* 12th Internat. Conf. of IBCC and EOAC., 14-18 sept 92, Noordwijkerhout, pp 285-290.

Kooijman, S.A.L.M. 1995. The stoichiometry of animal energetics. *J. Theor. Biol.* 177: 139-149.

Ratsak, C.H., Kooijman, S.A.L.M. & Kooi, B.W. 1993. Modelling of growth of an oligochaete on activated sludge. *Water Research* 27: 739-747

Ratsak, C.H., Kooi, B.W. & Kooijman, S.A.L.M. 1995. Modeling the individual growth of *Tetrahymena* sp. and its population consequences. *J. Euk. Microbiol.* 42: 268-276.

Stouthamer, A.H. & Kooijman, S.A.L.M. 1993 Why it pays for bacteria to delete disused DNA and to maintain megaplasmids. *A. van Leeuwenhoek* 63: 39-43.

Visser, J.A.G.M.de, Maat, A.ter & Zonneveld, C. 1994 Energy budgets and reproductive allocation in the simultaneous hermaphrodite pond snail, *Lymnaea stagnalis* (L.): A trade-off between male and female function. Am. Nat. 144: 861-867.

Zonneveld, C. & Kooijman, S.A.L.M. 1989. The application of a dynamic energy budget model to *Lymnaea stagnalis*. *Functional Ecology* 3: 269-278.

Zonneveld, C. & Kooijman, S.A.L.M. 1993. Comparative kinetics of embryo development. *Bull.*

Math. Biol. 3: 609-635.

Zonneveld, C. & Kooijman, S.A.L.M. 1993. Body temperature affects the shape of avian growth curves. *J. Biol. Syst.* 1: 363-374

Zonneveld, C. & Kooijman, S.A.L.M. 1989. Application of a general energy budget model to *Lymnaea stagnalis*. *Functional Ecology* 3: 269-278.

Zonneveld, C. 1992. Animal energy budgets: a dynamic approach. PhD-thesis, Vrije Universiteit dec 92.

Population level

Kooijman, S.A.L.M. 1986. Population dynamics on the basis of budgets. In: Metz, J.A.J. & Diekmann, O. (eds). *The dynamics of physiologically structured populations*: 266-297. Springer Lecture Notes in Biomathematics. Springer-Verlag, Berlin.

Kooijman, S.A.L.M. 1992. Biomass conversion at population level. In: DeAngelis, D.L. & Gross, L.J. (eds) *Individual based models; an approach to populations and communities.*: 338-358. Chapman & Hall.

Kooijman, S.A.L.M. 1994. Individual based population modelling. In: Grasman, J. & Straten, G. van (eds): *Predictability and Nonlinear Modelling in Natural Sciences and Economics*. Kluwer Academic Publishers, Dordrecht: 232-247.

Kooijman, S.A.L.M., Hoeven, N.van der & Werf, D.C.van der 1989. Population consequences of a physiological model for individuals. *Functional Ecology* 3: 325-336.

Kooijman, S.A.L.M. & Kooi, B.W. 1995. Catastrophic behaviour of myxamoebae. *Nonlin. World* 3: 77-83.

Kooijman, S.A.L.M., Kooi, B.W. & Boer, M.P. 1995. Rotifers do it with delay. The behaviour of reproducers vs dividers in chemostats. *Nonlin. World* 3: 107-128.

Kooijman, S.A.L.M., Muller, E.B. & Stouthamer, A.H. 1991. Microbial dynamics on the basis of individual budgets. *Antonie van Leeuwenhoek* 60: 159-174.

Kooi, B.W. & Boer, M.P. 1995. Discrete and continuous time population models, a comparison concerning proliferation by fission. J. Biol. Systems 3: 543-558.

Kooi, B.W. & Kooijman, S.A.L.M. 1994. Existence and stability of microbial prey–predator systems. *J. Theor. Biol* 170: 75-85

Kooi, B.W. & Kooijman, S.A.L.M. 1994. The transient behaviour of food chains in chemostats. *J. Theor. Biol* 170: 87-94

Kooi, B.W. & Kooijman, S.A.L.M. 1995. Many limiting behaviours in microbial food chains. In: Arino, O., Kimmel, M. & Axelrod, D. (eds) *Mathematical Population Dynamics.*, vol 2 Wuerz Publ, Winnipeg, Canada: 131-148.

Kooi, B.W. & Kooijman, S.A.L.M. 1995. Mass balance versus logistic equation in food chains. Proc. 4th Internat. Conf. Math. Pop. Dynamics, Houston. (to appear)

Kooi, B.W., Boer, M.P. & Kooijman, S.A.L.M. 1995. On the use of the logistic equation in food chains. (*Ecol. Mod.*, subm.)