The influence of design characteristics on statistical inference in non-linear estimation; a simulation study based on survival data and hazard modelling.

J.S. Andersen†, J.J.M. Bedaux‡, S.A.L.M. Kooijman‡, H. Holst†

†Department of Mathematical Modelling, Technical University of Denmark.

‡Department of Theoretical Biology, Free University of Amsterdam, The Netherlands.

Correspondence to Jens Strodl Andersen, IMM, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark, phone +45 45253376, fax +45 88452673, email jsa@imm.dtu.dk.

Abstract

[1] This paper describes the influence of design characteristics on the statistical inference for an ecotoxicological hazard-based model, using simulated survival data. The design characteristics of interest are the number and spacing of observations (counts) in time, the number and spacing of exposure concentrations (within c_{min} and c_{max}) and the initial number of individuals at time 0 in each concentration. A comparison of the coverage probabilities for confidence limits arising from the profile-likelihood approach and the Wald-based approach is carried out. The Wald-based approach is very sensitive to the choice of design characteristics whereas the profile-likelihood approach is more robust and unbiased. Special attention is paid to estimating a parametric no-effect concentration in realistic small-sample situations, since this is the most interesting parameter from an environmental protection point of view.

Keywords: large-sample theory, no-effect concentration, profile-likelihood, small samples.

1 Introduction

Evaluation of environmental or human risks posed by potentially hazardous pollutants or chemicals, requires results from chemical analyses and results from toxicity tests. During the last decade, toxicity tests have played an increasingly important role in hazard identification and in calculation of toxicity measures as basis for the estimation of a safe level for a substance in the environment or the estimation of an acceptable daily intake. The estimations involve safety factors to include inter- to intra-species, acute-to-chronic and laboratory-to-field extrapolations (Chapman *et al.* 1998). A solid basis for such extrapolations requires a wide-ranging test battery. Survival is one factor evaluated in the ecotoxicological test battery (others are inhibition of growth and reproduction, and mutation).

Possible endpoints of these toxicity tests are NOEC's (No Observed Effect Concentrations), EC_x-values (Concentration showing x% Effect) or NEC's (No Effect Concentration). The NOEC has been severely criticized on statistical grounds by several authors (Pack 1993, Suter 1996). The main objection is that the calculation of the NOEC is based on a statistical test, and statistical tests are not designed to estimate a parameter. This methodological problem has several negative consequences:

[1]**This paper has appeared in 2000 in the Journal of Agricultural, Biological and Environmental Statistics, Volume 5, Pages 323-341.**

- the NOEC is not a safe concentration (non-significant does not mean non-existent)
- it is always one of the toxicity-test concentrations
- nothing can be derived about its precision
- only part of the data is being used
- the value of the NOEC depends on the chosen statistical test

As an alternative EC_x-values have been proposed by Pack(1993) and NEC's by Kooijman(1993). In several workshops organized by SETAC and OECD it was concluded and recommended that the NOEC should be abandoned, that regression-based estimation procedures should be used and that time should be incorporated in the analytical procedures for toxicity tests (OECD, 1998). More research was required to validate threshold and time to response models and to find optimal experimental designs (Chapman *et al.* 1996). Apart from the statistical debate about the NOEC there is a continuing debate about the existence of thresholds in response curves of biological systems (Cairns, 1992). According to Van Straalen (1997) an answer to this problem cannot be given without a mechanistic interpretation of the concentration-response relationship. The approach of Kooijman and Bedaux (1996) is regression based and incorporates time of exposure. It is also mechanistic, based on biological assumptions.
This paper aims to evaluate the effects of experimental design on statistical inference concerning the ecotoxicological survival model of Bedaux and Kooijman (1994). The evaluation is based on Monte-Carlo simulations and is focused on statistical properties of the NEC.
In aquatic ecotoxicological survival experiments the individuals are living in a well defined medium containing a range of concentrations of the test compound. The physico-chemical environment is kept constant during the test period and the exposed units are optimally fully randomised to treatment conditions to eliminate external effects. Each individual is only exposed to one concentration. The numbers of surviving organisms are registered at the end of the experiment, but preferably also at several intermediate time-points. An example of a data-set from a toxicity survival experiment is shown in table 1. It concerns the effect of the pesticide dieldrin on guppies (mean length 4.5 cm) at 15° C. The data clearly show that toxic effects are time dependent: after one day only the high concentrations show a large effect, after 7 days also the intermediate concentrations show large effects.
The experimental design characteristics of interest, in this study, are: the number and spacing of observation times t_i, the number and spacing of exposure concentrations c_j (within c_{min} and c_{max}) and the initial number of individuals x_{0j}, at time 0 in each concentration c_j. The simulation study was designed to look at the influence of the above mentioned characteristics on the following issues:

- The size and shape of the confidence sets of the estimated parameters. The profile-likelihood approach is compared with the standard Wald-based approach.
- The extreme values or outliers of parameter estimates.
- The kinetic type (explained in the model description).
- The censoring of time of death of each individual.

Table 1: Number of surviving guppies *Poecilia reticulata* in natural sea water after exposure to the pesticide dieldrin. Data from IMW-TNO Laboratories, Delft.

	concentration *dieldrin* (μg l^{-1})							
time (d)	0	3.2	5.6	10	18	32	56	100
0	20	20	20	20	20	20	20	20
1	20	20	20	20	18	18	17	5
2	20	20	19	17	15	9	6	0
3	20	20	19	15	9	2	1	0
4	20	20	19	14	4	1	0	0
5	20	20	18	12	4	0	0	0
6	20	19	18	9	3	0	0	0
7	20	18	18	8	2	0	0	0

Datasets like the one presented above are typical in toxicity assessment of environmental samples and of new chemicals. Very little is known about how the accuracy of the methods used for making inference about the endpoints (i.e. no-effect concentrations) in such experiments depends on the experimental design. Chapman *et al.* (1996) recommend that research is required into the effect of statistical accuracy and precision of the number and spacing of concentrations. The purpose of the current simulation study is to explore the dependency between experimental design and inference made about the endpoints, using simulated survival data.

2 THE SURVIVAL MODEL

2.1 BIOLOGICAL BACKGROUND AND HAZARD MODELING

The survival model we investigate is based on simple biological assumptions (Bedaux and Kooijman 1994). The central assumption is that the hazard rate (defined as the instantaneous probability, per time increment, that death strikes at a certain age given survival up to that age) is proportional to the concentration of the chemical compound in the animal as far as it exceeds the internal no-effect level Q_0. The uptake dynamics are described by a one compartment model involving the uptake rate k_u and elimination rate k_e.
The concentration in the test environment is considered constant over time and is denoted c. The initial concentration of the compound in the animal, is considered negligible.
This leads to

$$Q(t;c) = \frac{k_u}{k_e} c(1 - \exp(-k_e t)) \ . \tag{1}$$

The environmental no-effect concentration c_0 is defined as the highest value of c at which $Q(t;c)$ does not exceed Q_0 even after long exposure. It can be easily seen that $c_0 = Q_0 k_e / k_u$.
Now we can write

$$Q(t;c) - Q_0 = \frac{k_u}{k_e} \left[c(1 - \exp(-k_e t)) - c_0 \right] \ .$$

If $c > c_0$ then and $Q(t;c)$ will exceed Q_0 at $t_0 = (-1/k_e)\ln(1 - c_0/c)$. The hazard rate $h(t;c)$ is by assumption proportional to $Q(t;c) - Q_0$ and if we include a background mortality rate,

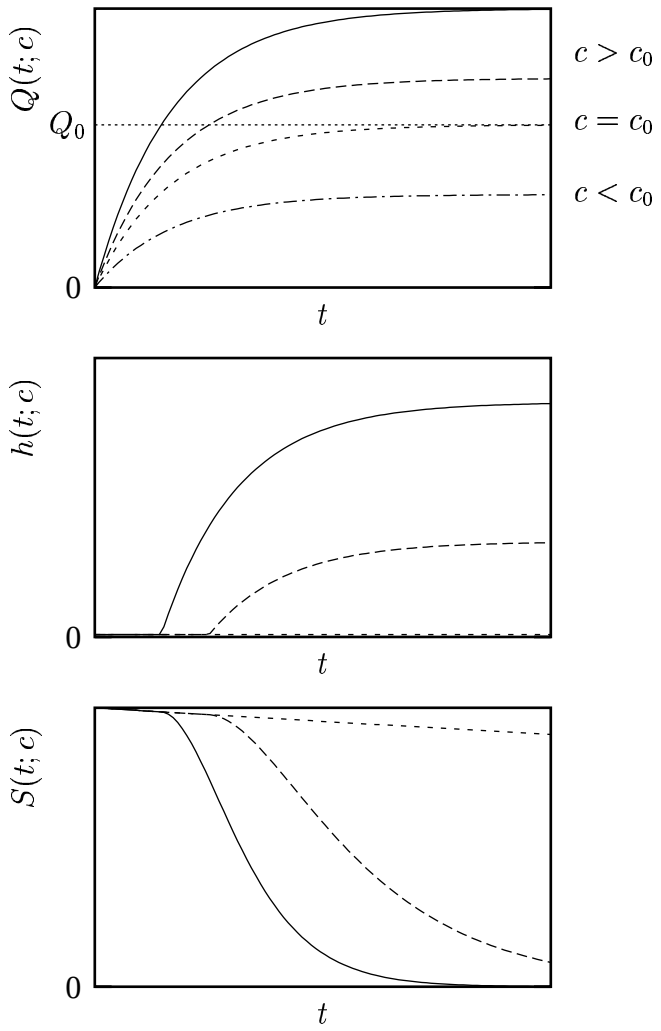

Figure 1: Internal concentrations for various values of c and the related hazard functions and survivor functions (including background mortality).

the total hazard rate can now be written as

$$h(t;c) = \lambda + k_\dagger \left(c(1 - \exp(-k_e t)) - c_0\right)_+ \tag{2}$$

where λ is the background mortality rate (assumed constant), and $(x)_+$ refers to the maximum of x and 0. A constant background mortality is a reasonable assumption, because the duration of the experiment is short compared with the mean life time of the organisms used. The parameter $k_\dagger$ is the product of the bioconcentration factor k_u/k_e and the proportionality constant between the internal concentration and the (toxic) hazard rate. It is called the killing rate and it expresses the degree of toxicity above the no-effect concentration with respect to survival. Figure 1 shows curves of the internal concentration Q, the hazard rate h and the survivor function S for various values of c.

Parameter	Symbol	Estimate	Units	Std. Dev.
Blank mortality rate	λ	0.00835	d^{-1}	0.00490
No-effect concentration	c_0	5.20	$\mu g l^{-1}$	0.465
Killing rate	$k_\dagger$	0.0376	$l\mu g^{-1}d^{-1}$	0.00777
Elimination rate	k_e	0.791	d^{-1}	0.281
Deviance		36.43		
95% Confidence interval:				
Wald-based approach	c_0	[4.29:5.30]		
Profile Likelihood approach	c_0	[2.72:2.84] $\cup$ [4.09:6.92]		

Table 2: Results of Maximum Likelihood estimation on the data in table 1, note that separate confidence intervals are possible using the profile likelihood approach.

2.2 SURVIVAL MODEL

Though above described assumptions are fairly simple in a biological context, the survivor function $S(t;c)$ of time of dying (defined as the probability that a random variable exceeds a specific value) is non-linear and contains four parameters.

$$S(t;c) = \begin{cases} \exp\left(\frac{k_\dagger}{k_e}c(e^{-k_e t_0} - e^{-k_e t}) - k_\dagger(c-c_0)(t-t_0) - \lambda t\right) & \text{if } c > c_0 \text{ and } t > t_0 \\ e^{-\lambda t} & \text{otherwise} \end{cases} \quad (3)$$

The four parameters are the no-effect concentration c_0, the killing rate $k_\dagger$, the elimination rate k_e and the background mortality rate λ (assumed constant). Model (3) is referred to as "normal kinetics".
The maximum likelihood estimates of the model parameters for the data presented in table 1 are shown in table 2 together with the 95% confidence intervals arising from the Wald-based approach and the profile likelihood based approach. In figure 2 the data are shown together with the fitted model curves.

2.3 SPECIAL SITUATIONS

In two important special situations, when the elimination rate k_e becomes very small ($k_e \to 0$) or very large ($k_e \to \infty$), the kinetics is referred to as "slow kinetics" or "fast kinetics". If $k_e \to 0$, then $c_0 \to 0$ and $k_\dagger \to \infty$, and model (3) reduces to

$$S(t;c) = \exp\left(-\frac{1}{2}\kappa_\dagger c((t - \frac{x_0}{c})_+)^2 - \lambda t\right). \quad (4)$$

The three parameters are the killing acceleration $\kappa_\dagger$, given by $\kappa_\dagger = \lim k_\dagger k_e$, the background mortality rate λ and the compound parameter $x_0 = c_0/k_e$. The ratio x_0/c equals t_0, the time at which the internal no-effect concentration is exceeded. The name 'killing acceleration' reflects both the roots and the time dimension of this parameter: the killing acceleration basically derives from the killing rate, the essential difference being in the dimensions of these two parameters.
When $k_e \to \infty$ model (3) reduces simply to

$$S(t;c) = \exp\left(-k_\dagger t(c-c_o)_+ - \lambda t\right). \quad (5)$$

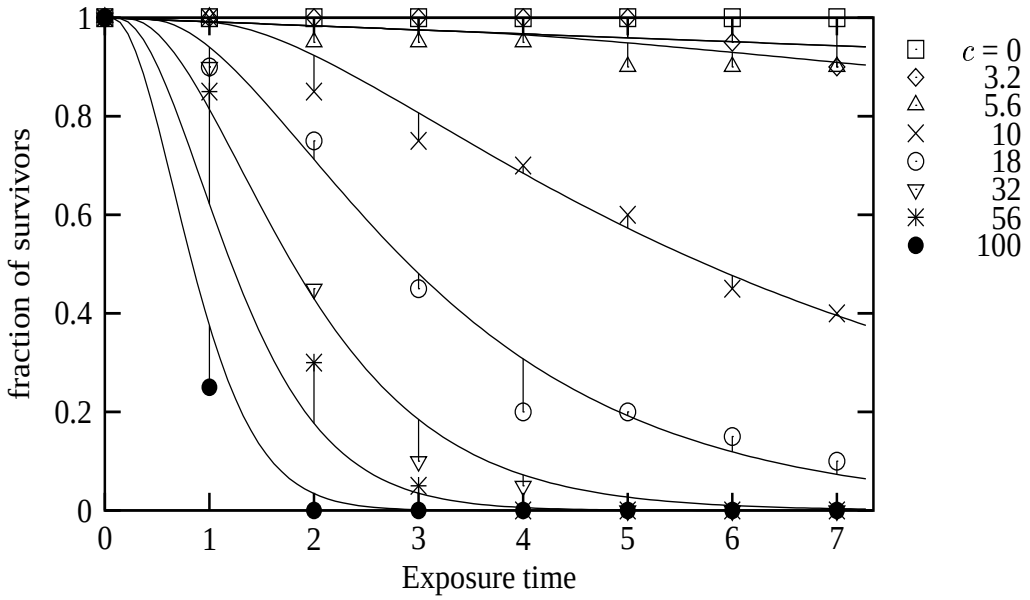

Figure 2: Empirical data from table 1 with estimated survivor curves.

2.4 Statistical Aspects of Experiments and Model Structure

Several features of the model lead to potential difficulties. The importance of these problems will be addressed using simulations. The potential difficulties are:

1. The experimental set-up induces right intervally censoring due to the fixed length of the experiment. The fixed observation time-points reduces the information of the exact time of death between consecutive observation times.

2. The estimator of λ is a discrete variable if the number of individuals per concentration below the c_0 is small.

3. The no-effect concentration makes the survivor function non-differentiable in the points $(c, t_0(c))$. This may affect the asymptotic properties of the maximum likelihood estimates.

4. The killing rate in fact is a compound parameter, equal to the product of the bioconcentration factor k_u/k_e and the proportionality constant between hazard and internal concentration. This generates a dependence between the elimination rate and the killing rate. Further the fraction $k_\dagger/k_e$ occurs in the survivor function. The dependence might influence the asymptotic properties of the estimated parameters.

3 Estimation of the parameters

The experimental response is the number of surviving organisms x_{ij} at a fixed time t_i, $i = 0, 1, \ldots, r$ exposed to the concentrations $c_j, j = 1, \ldots, k$, where $c_1 = 0$, the control condition. The probability p_{ij} that an organism, exposed to the concentration c_j, will die between t_{i-1} and t_i is given by $p_{ij} = q(t_{i-1}, c_j) - q(t_i, c_j)$. The number of organisms n_{ij} which died in that period is given by $n_{ij} = x_{i-1,j} - x_{ij}$. The number of organisms surviving at t_r will be denoted by $n_{r+1,j}$ and the probability of surviving at t_r is denoted by $p_{r+1,j}$ and equals $S(t_r, c_j)$.

The probability that the number of surviving organisms is x_{ij} can now be written as a product of multinomial probabilities:

$$Prob(\underline{x}_{ij} = x_{ij}) = Prob(\underline{n}_{ij} = n_{ij}) = \prod_{j=1}^{k} x_{0j}! \prod_{i=1}^{r+1} \frac{p_{ij}^{n_{ij}}}{n_{ij}!} \ . \tag{6}$$

Parameter estimation based on the simulated data is done using the maximum likelihood method. The log-likelihood function is given by

$$l(\theta; (x_{ij})) = \sum_{i=1}^{r+1} \sum_{j=1}^{k} n_{ij} \ln(p_{ij}), \quad \text{with } \theta = (c_0, k_\dagger, k_e, \lambda) \tag{7}$$

where the constant term only containing multinomial coefficients has been ignored. Maximum likelihood estimates can be found by solving the vector equations $G(\theta) = \partial l / \partial \theta = 0$ (McCullagh and Nelder 1989).
As a rough measure of goodness-of-fit the deviance of the model is used. The deviance is defined as twice the difference between the maximum achievable log likelihood and that attained under the fitted model. The maximum achievable log likelihood $l_{sup}(x_{ij})$ is obtained by estimating each p_{ij} without any constraints, i.e., $\hat{p}_{ij} = n_{ij}/x_{0j}$ and substituting this into (7).
The difference in deviances of nested models can be used as a test statistic in the likelihood-ratio test. The distribution of the test statistic is asymptotically χ^2_d distributed where d is the difference between the number of parameters of the two nested models. The deviance should not be used to test the absolute goodness-of-fit, since the asymptotic theory does not always apply because many of the expected counts $E(\underline{n}_{ij})$ are too small.

4 Inference about the model parameters

The confidence limits for the estimated parameters are calculated based on the likelihood ratio interval, described as the profile likelihood by Williams (1986) and Aitken (1986). If $L(\theta|X)$ denotes the likelihood function, then the profile likelihood of θ_i, $P(\theta_i)$ is definded as

$$P(\theta_i) = \mathrm{Max} L(\theta|X, \theta_i) \tag{8}$$

or, in words, $P(\theta_i)$ is obtained by fixing the parameter θ_i and estimating all other parameters. The $100(1-\alpha)\%$ profile likelihood interval for θ_i is given by the solution to

$$2 \log \left[P(\hat{\theta}_i) / P(\theta_i) \right] = \chi^2_1(1-\alpha) \tag{9}$$

where $\chi^2_1(1-\alpha)$ is a percentile of the chi-square distribution with one degree of freedom.
The $100(1-\alpha)\%$ asymptotic Wald-based confidence interval is based on the asymptotic linearization result $\hat{\theta} \sim N(\theta, D(\hat{\theta}))$, which holds under appropriate regularity conditions, (Seber 1989). The variance-covariance matrix of the maximum likelihood estimate $D(\hat{\theta})$ can be calculated as the inverse of the information matrix $I(\hat{\theta})$ (defined as minus the expectation of the matrix of the second derivatives (McCullagh and Nelder 1989)). For large n the $100(1-\alpha)$ % interval for θ_i is given by

$$\hat{\theta}_i \pm t^{\alpha/2}_{n-4} \sqrt{D_{diag}(\hat{\theta}_i)} \tag{10}$$

where t_{n-4} is the t-distribution with $n-4$ degrees of freedom. The t-distribution is used in combination with asymptotic theory since it is common practice.

Several measures can be applied to compare the two methods of calculating confidence sets. One intuitively reasonable way is to describe their coverage probabilities, relative to the true parameter values, as a function of design characteristics. A second is to compare the size of the confidence sets conditioned on the coverage probabilities as a rough measure of precision. A third way, of special interest from an environmental perspective, is whether the estimated c_0 is over- or under-estimated. It is critical if the true value of the no-effect concentration is below the lower estimated confidence limit (over-estimation). This is of practical interest because we want to avoid the situations that we judge an environmental risk to be small while it is actually large. Further it is of practical interest to formulate a criterion on which the reliability of the different confidence sets can be judged, i.e. when can we trust the asymptotic large sample intervals, the profile likelihood intervals and when are these intervals suspect?
One way of evaluating the assumptions for the profile likelihood estimation of the confidence sets is by comparing the deviance from the maximum likelihood estimation with the deviance obtained by fixing an estimated parameter. The difference should asymptotically follow a $\chi^2(1)$ distribution. By calculating the tail (inverse) probability (of the differences in the deviances from the full and the reduced model), and then the c.d.f. of these, the result should be asymptotically uniform distributed.
The estimation of the parameters, the estimation of the profile-likelihood and the calculation of the large-sample parameter variance were performed using the DEBtox program (Kooijman and Bedaux 1996). The processing of these results and the graphics were performed using S-PLUS (S-PLUS1995).

5 Monte-Carlo simulations

5.1 GENERATION OF DATA.

Given the model, the chosen parameter values, the chosen time points and the chosen concentration levels, the cumulated probability of dying can be calculated for each concentration-time point (using equation (3)). The initial number of individuals at each concentration level is chosen and for each individual a random number is generated (between 0 and 1, using a slightly modified and adapted version of RAN1 (Press *et al. 1992*)). This random number is compared with the cumulated probabilities belonging to the concentration. When the random number exceeds the cumulated probability at a concentration / time point the individual is considered dead at that time point. This procedure is used to obtain an observation matrix such as table 1, and to arrive at parameter estimates. These estimates are then compared with the chosen values.

5.2 DESIGN CHARACTERISTICS.

It is common practice for aquatic ecotoxicity survival tests to have 5 to 7 concentrations plus 1 control, i.e. k ranges from 6 to 8. In this study the range of concentrations is a bit wider: k ranges from 5 to 10. The nonzero-concentrations are spaced equidistantly on a log scale. The simulations are divided into several scenarios, each scenario was designed to explore the issues listed in the introduction.
The first three simulation scenarios were carried out with the following parameter values ($c_0 = 2.0$, $k_\dagger = 0.05$, $k_e = 0.5$ and λ=0.01) and time-points $(0, 1 \ldots, 7)$. The choice for the parameter values was based on the results in table 2.

Scenario 1: 1000 simulations with each combination of (k = 4 to 9 concentrations in exponential series, as 0, $100^{(j-1)/(k-1)}$ for $j = 1, \ldots, k-1$) and (initial number of individuals per concentration, $x_{0j} = [5, 6, 7, 9, 11, 15, 25, 35, 50, 75, 100]$).
Scenario 2: If a number of individuals $x_{0_{total}}$ is available for a toxicity test, several experimental designs can be chosen by varying k, the number of concentratons, and x_{0j}, the number of individuals per concentration. To study the effect of this choice, $x_{0_{total}}$ was varied in the range [40;100] and all possible combinations of k and x_{0j} were investigated (with the minimum limits of $k = 4$ and $x_{0j} = 4$). The concentrations were, as in scenario 1, chosen in exponential series.
Scenario 3 To study the effect of censoring in time, 1000 simulations were made and each simulation results in 3 data-sets with decreasing information in time. The first dataset contains the exact time of death of all the individuals. This means full information. The second contains the exact time of death until time-point 7, i.e. the first dataset censored at time point 7. The third contains a discrete version of the second dataset – the summed number of dead individuals in the time intervals 0 to 1, 1 to 2, . . . , 6 to 7.
Scenario 4 The effect of concentrations below c_0, was analysed by making 5000 simulations with the parameter values ($c_0 = 10.0$, $k_\dagger = 0.05$, $k_e = 0.5$ and λ=0.01), time-points $(0, 1 \ldots, 7)$, $x_{0j} = 50$, and concentrations (($c^*, 12.1, 14.5, 19.5, 28.2, 42.5, 100$), where c^* in turn equals the four values: $\{10.0, 6.67, 3.33, 0.0\}$)
The reproducibility and simulation size were checked by making additional simulations with the same k, x_{0j} combinations and additional simulations with N_{sim} =2000, . . . , 4000.

6 Results

6.1 KINETICS

Before discussing the coverage probabilities for the profile likelihood and large sample approach a short comment is needed concerning the estimated kinetic type. The observation matrix is generated using normal kinetics (equation 3), but the estimation procedure is free to choose the kinetic type resulting in the smallest $\log$ likelihood. If slow kinetics is chosen, most of the kinetic parameters have a new meaning, another numerical value and unit and therefore no direct comparison is possible for all the parameters. The percentage of estimations resulting in slow kinetics is shown in figure 3. No slow kinetics occur, when the total number of individuals exceeds 100, therefore these combinations (of c_j and x_{0j}) are omitted from the figure. It is evident that the percentage is decreasing with the increasing number of concentrations and individuals. This is due to a better determination of the kinetic parameters k_e and $k_\dagger$. The kinetic parameters describe the curvature of the survivor function surface. The iso-animal lines in figure 3 indicate that less slow kinetics occur when using a higher number of concentrations with fewer animals per concentration. This gives more points on the surface and the curvature is better described.
The percentage of slow kinetics is influencing the coverage probabilities of the different parameters in the following way:

6.1.1 Large-Sample Inference

For slow kinetics the parameters c_0, k_e and their standard deviations are by definition identically 0. This is because the concentration in the animal builds up without limit, so we are sure that it exceeds the actual no-effect concentration eventually. Thus $k_\dagger$ becomes $\kappa_\dagger$. The parameter λ

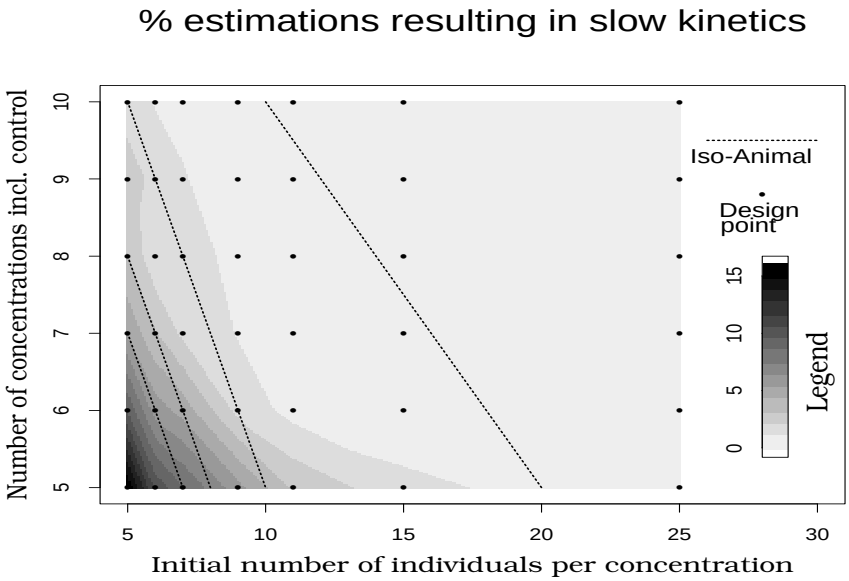

Figure 3: Probability of slow kinetics (in %) - no slow kinetics occur for more than 25 initial individuals per concentration.

is not influenced. The adjustment for slow kinetics for c_0, k_e and $k_\dagger$ are calculated by assuming that the percentage of true parameters values outside the confidence limits are the same for slow and normal kinetics. This method is preferred over discarding the datasets from the analysis. The assumption is thought to favour the unbiasedness of the Wald-based approach compared to the profile likelihood based approach.

6.1.2 Profile Inference

In case of slow kinetics the point estimate of c_0 will be zero. Then, for physical reasons, the lower confidence bound for c_0 is set identical to 0. The upper bound can still be calculated, so there is no need for an adjustment.

6.2 RESULTS FROM SCENARIO 1

Bias, 95% Confidence limits, Large Sample. If the results are unbiased we expect θ_i to be in the 95%-confidence interval (cfi.) in 95% of the cases. The coverage probabilities, adjusted for slow kinetics of the nominal 95% confidence limits for the large sample approach regarding the four parameters are shown in figure 4.
The estimation of the background mortality rate λ is strongly dependent on the number of animals used as the percentage outside the cfi. ranges from 5 to 50%. The bias is large for small sample sizes. There is a small effect due to the distribution of the animals; many animals per concentration improves the coverage probability.
The estimation of the elimination rate k_e is not as sensitive as λ. The percentage outside the cfi. ranges from 5 to 13%, and the effect of animal distribution is the opposite; many concentrations improve the coverage probability. The surface appears complex with no single maximum.
The estimation of the killing rate $k_\dagger$ is the least sensitive, the percentage outside the confidence interval ranges from 4 to 8%, and no clear effect of animal distribution is present.
For the no effect concentration c_0 , the percentage outside the cfi. ranges from 5 to 25%, and no clear effect of the animal distribution can be observed.

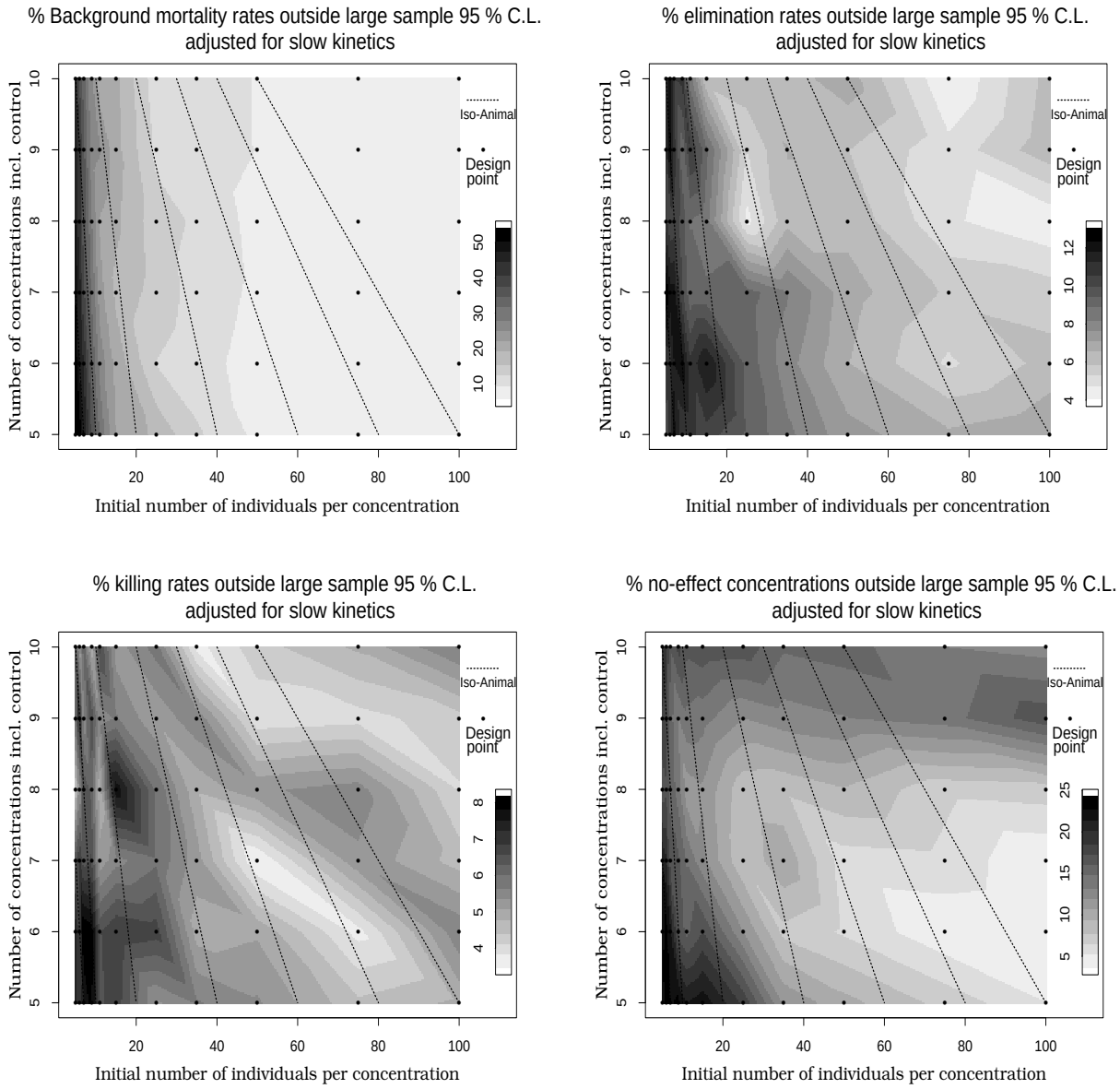

Figure 4: Coverage probabilities for 95% large sample confidence intervals

When the number of animals and concentrations is decreasing the number of extreme parameter values and outliers is increasing.
The reason for the large bias of the large sample confidence limits, for small sample sizes, could be due to several causes, including

- The skewness of the simulated parameter results (small samples, illustrated in figure 5)
- The non-differentiability of $S(t;c)$ in the points $(t_0(c), c)$.
- The right censoring in time of the observations.
- The second order symmetric Taylor approximation of the likelihood function is too crude.
- The influence of the correlation between $k_\dagger$ and k_e, as illustrated in figure 5.
- The discreteness of λ for small initial number of animals, as illustrated in figure 5.

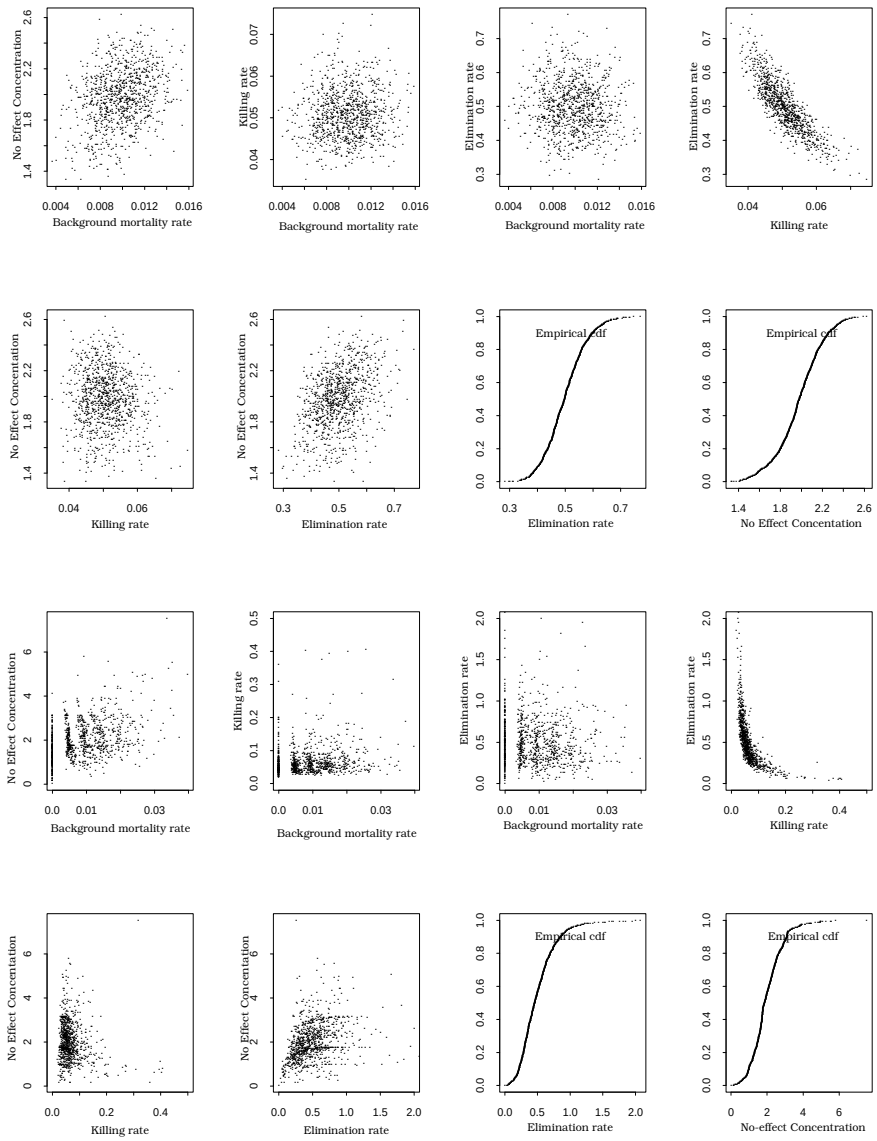

Figure 5: The upper 8 plots are $k = 10$ concentration groups with $x_{0j} = 100$ animals per concentration group, showing the large sample situation. The lower 8 plots are $k = 10$ and $x_{0j} = 9$, showing the small sample situation. The plots illustrate the violated assumptions, that influence the asymptotics, when sample sizes decrease.

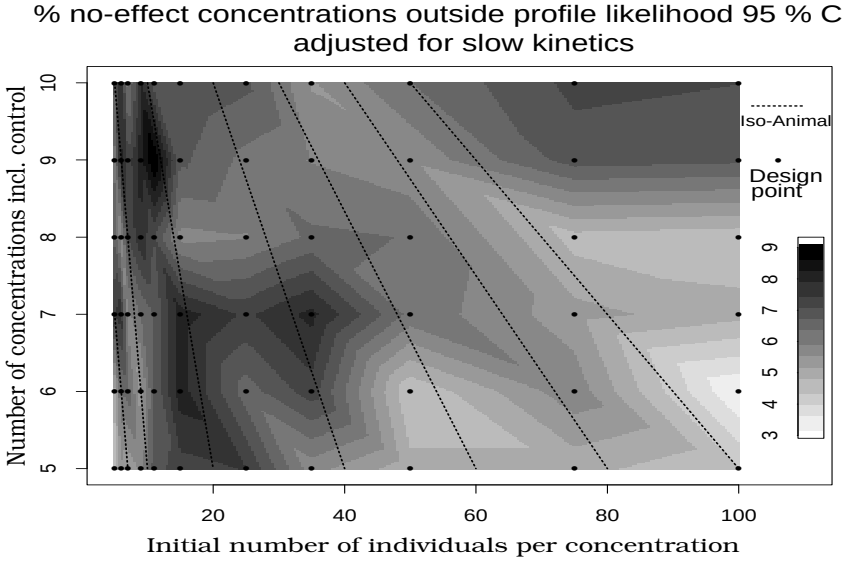

Figure 6: Coverage probabilities for 95% profile likelihood confidence intervals

Bias, 95% Confidence limits, profile likelihood. Ecotoxicological tests are usually performed with a legislative purpose or a classification purpose. Therefore the most interesting parameter, in that content, is the no-effect concentration c_0 and the profile likelihood interval is only calculated for the estimated c_0 values. The coverage probability is shown in figure 6.
The percentage outside the cfi. ranges from 3 to 9%. There is only a small effect due to sample size and no effect due to the distribution of the animals. The contour plot suggests a complicated surface.
Precision. The dependency between design characteristics and precision of the estimated parameters is illustrated in figure 7. The precision is expressed as: the mean of the estimated parameter minus the lower 95 % profile confidence limit conditioned on normal kinetics and that the estimated parameter is included in the interval: $E(c_0 - Low_{95} | (\text{normal kinetics}), c_0 \in 95\%)$. As expected the precision increases with increasing number of concentrations and number of individuals. This pattern is, as expected, the same for all parameters using the large sample approach.
Oveestimation/Underestimation. To evaluate the over or under-estimation of the c_0, the proportion of the estimated c_0 values outside the confidence region was divided by the proportion below the lower limit. As stated above, from an environmental protection point of view, it is critical if the ratio is below 2. It should be noted that nothing in the estimation theory states that the profile likelihood interval, as opposed to the large sample interval should be symmetric. In general there is no design-dependent over- or underestimation for the profile likelihood approach, but a very large overestimation for the large sample approach increasing as the sample size decreases, as illustrated in figure 8. This result is also valid for λ, k_e and $k_\dagger$. Adjustment for slow kinetics is done as described previously, by assuming that the percentage of true parameter values outside / below the confidence limits are the same for slow and normal kinetics

6.3 RESULTS FROM SCENARIO 2

The aim of scenario 2 is to describe how the estimation depends on the distribution of a given number of individuals, i.e. 54 individuals can be assigned to 6 concentrations with 9 individuals

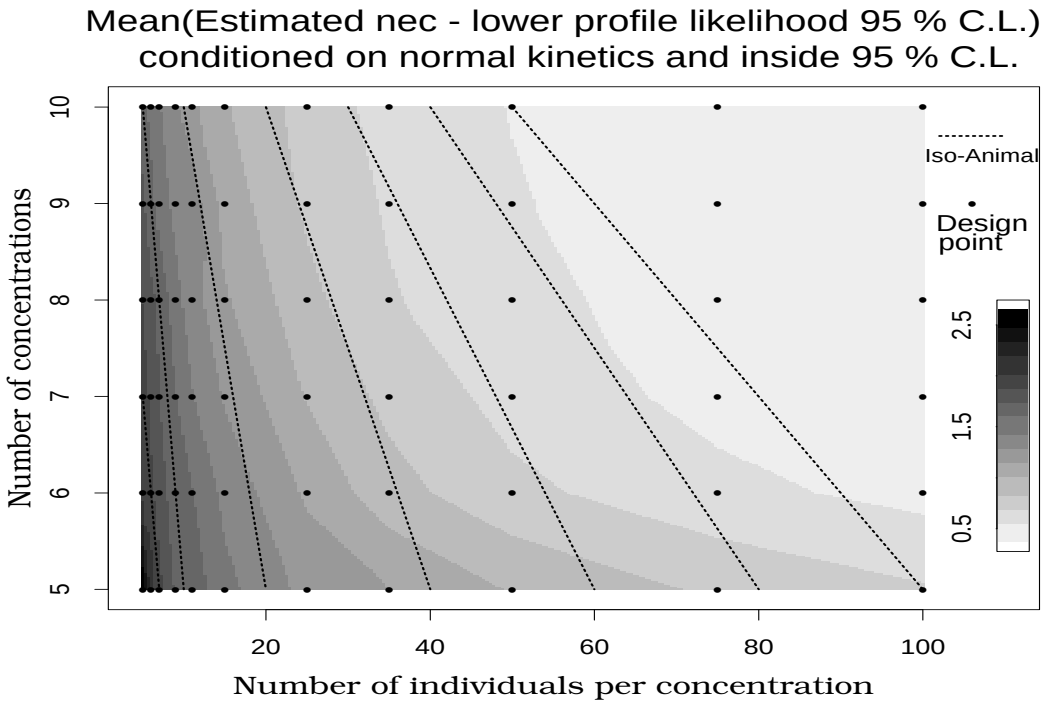

Figure 7: Precision expressed as: E($c_0 - Low_{95}|$norm, $c_0 \in 95\%$)

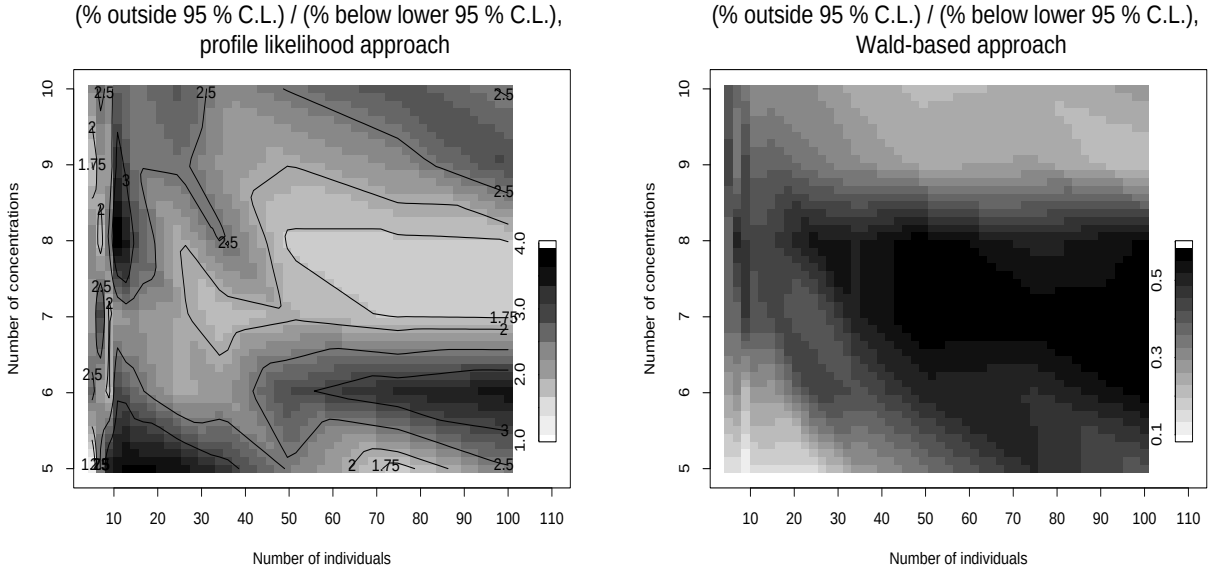

Figure 8: Overestimation/Underestimation of c_0: Percent true parameter values outside the 95% confidence interval divided by percent true parameter values below the 95% confidence interval

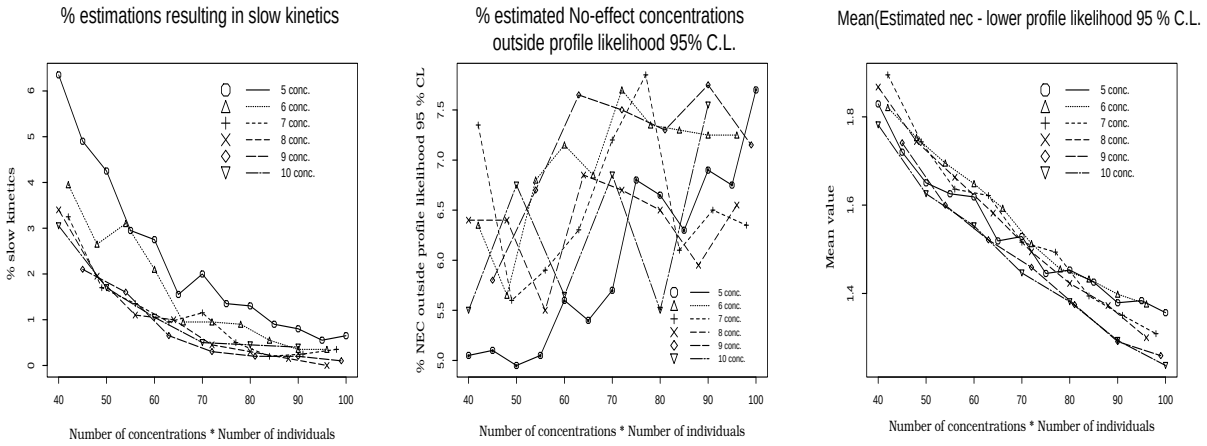

Figure 9: Results from scenario 2: Percent estimations resulting in slow kinetics, percent c_0 inside the estimated 95% confidence limits, and precision

per concentration, or to 9 concentrations with 6 individuals per concentration etc. The minimum and maximum number of animals used were 40 and 100 respectively. The large sample confidence intervals perform poorly so only the profile likelihood for c_0 is considered.

Kinetics. The probability of ending in slow kinetics (figure 9) depends on the design characteristics and in order to end up with normal kinetics the number of concentrations should be increased. The increased number of concentrations gives more points describing the curvature of the response surface and thereby the kinetic parameters. See also the remarks about slow kinetics in scenario 1.

Bias, 95% Profile-Likelihood Confidence Interval. The coverage probabilities of the 95% profile confidence limits regarding the c_0-parameter is calculated and the results presented in figure 9. The results from the previous scenario 1 concerning the robustness of the profile likelihood intervals and the general level of coverage is confirmed here and there is no clear dependency of design characteristics.

Precision. There is no evidence that the design has influence on the precision, though the precision is increased with an increasing number of individuals.

6.4 RESULTS FROM SCENARIO 3

Two major results can be derived from this "time to event" study:

• It is efficient to count the number of survivors in fixed time intervals. No substantial information is gained by observing the batch around the clock and thereby obtaining information about the exact time of death. This is concluded based on the upper nine plots in figure 10, where no significant difference is observed between the profile likelihood for the dataset with 7 discrete time-points and the exact time of death in 7 time-points.

• It is however important to continue the experiment until something has happened in the lower concentrations since substantial information is gained. Large changes are observed in the point estimate and in the profile likelihood function as the number of time points is increased (lower three plots in figure 10).

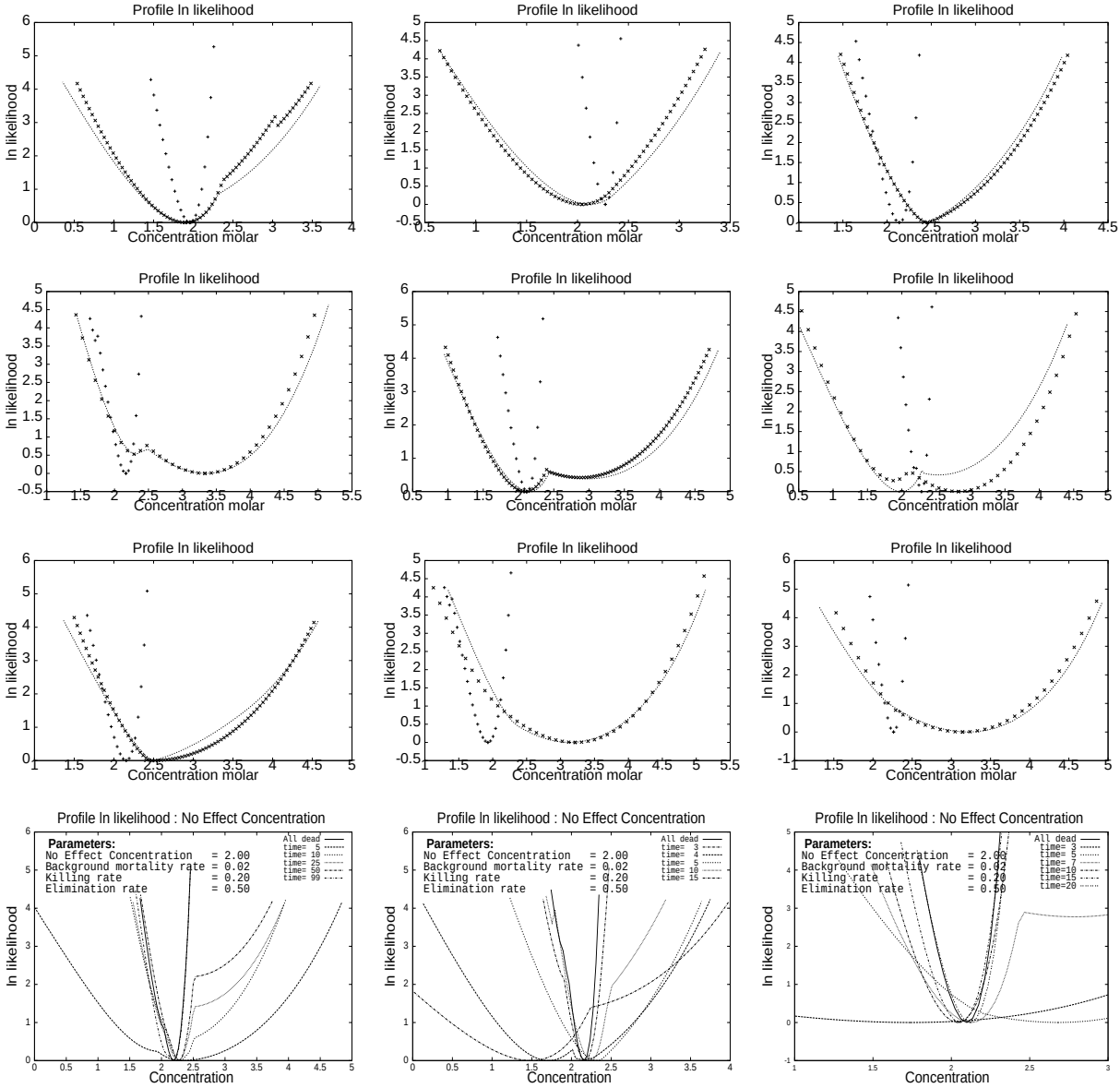

Figure 10: The legend for the upper nine c_0 profile likelihood plots: **x** is the exact time of death in the time interval [0,7]; **+** is exact time of death in the time interval [0;all animals are dead] and $\cdots$ is the cumulated deaths in the discrete time intervals $(0,1],(1,2],\ldots,(6,7]$. The lower three plots show the profile likelihood for c_0, and the information gained by continuing over time for discrete observation intervals $(0,1],(1,2],\ldots,(6,7]$. In all twelve plots $c_0 = 2$.

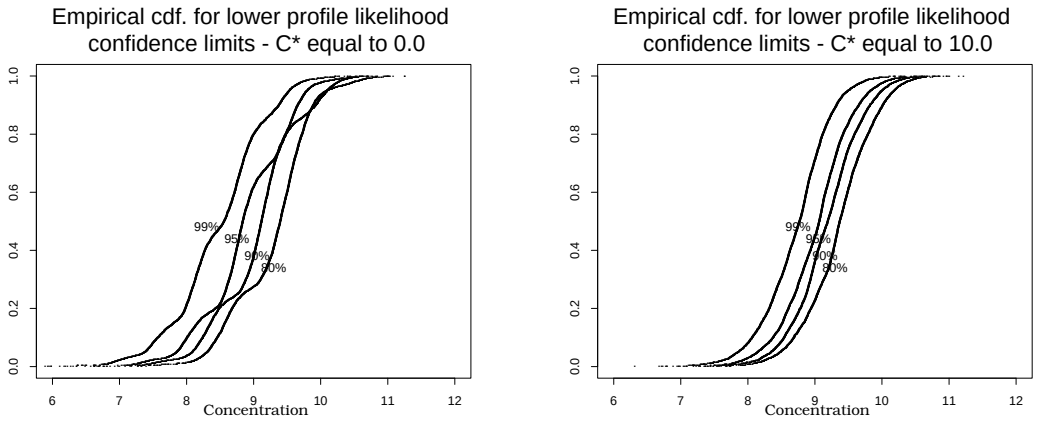

Figure 11: Effect on the profile likelihood, derived from the location of concentrations below c_0. (a) corresponds to $c* = 0$ and (b) corresponds to $c* = 10$ (the c_0-value)

6.5 RESULTS FROM SCENARIO 4

The location of the concentrations below c_0 are not influencing the mean point-estimate of the no-effect concentration or the empirical variance of the 5000 re-estimated parameter values. The location influences the profile likelihood and thereby the confidence sets. In figure 11, empirical cdf's for the lower profile likelihood confidence limits are shown. If the lowest concentration is far from c_0, the confidence band becomes more irregular.

7 Discussion and recommendations

The presented simulation study has favoured a large variation in the design characteristics rather than the examination of a large set of parameter values. The robustness of the conclusions outlined in the following should therefore not be looked upon as proofs.

It seems, based on the current study, that the coverage probability of the estimated confidence limits for the c_0 parameter is, for small sample sizes, improved using the profile likelihood method compared with the large sample approach. It is assumed that comparable improvements could be achieved for the other parameters. Therefore the profile likelihood method is recommended when sample sizes are moderate or small. This is in good agreement with the theoretical results discussed in Morgan (1996).

The accuracy of the estimated parameters depends on the experimental design. The optimal design for estimating a given set of parameters depends on the nature of the parameter, whether it describes a level or curvature. In this paper the trade-off is between the accuracy of $\hat{\lambda}, \hat{c}_0$ and $\hat{k}_\dagger, \hat{k_e}$ since a large initial number of animals per concentration favour estimation of λ, c_0, opposed to many concentrations that favour estimation of $k_\dagger, k_e$. If any of these factors are sparsely chosen, then a compensation by continuing the experiment in time is possible. This may lead to problems with the physico-chemical environment, aging, etc. Then a constant background mortality will not be acceptable any more.

If c_0 is the main parameter of interest the experiment needs to be prolonged until the lower concentrations have caused an effect. Otherwise the surface and its cutoff point c_0 is determined by- or extrapolated from mortality at high concentrations and short-time exposure while the aim

of the study is the estimation of c_0 after a long exposure.
It is only recommended to use resources to count the number of dead animals e.g. once a day since no substantial information is gained by reporting the exact time of death. The observation time points must off course be related to the kinetics of the tested compound or environmental sample.

Acknowledgements

This work was funded by the Danish Council of Technical and Scientific Research, through The Groundwater Research Centre, Technical University of Denmark.
We thank two anonymous referees and the editors for improvements and most helpful suggestions.

References

[1] **Aitken M.** 1986. Statistical modelling: The likelihood approach. *The Statistician,* **35**, 102-13.

[2] **Bedaux J.J.M.** and **Kooijman S.A.L.M.** 1994. Statistical analysis of bioassays, based on hazard modelling. *Environmental and Ecological Statistics*, **1**, 303-314.

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

[4] **Chapman P.F., Crane M., Wiles J.A, Noppert F.** and **McIndoe E.C.**(eds) 1996. Asking the right questions: Ecotoxicology and Statistics. SETAC-Europe, Brussels.

[5] **Kooijman S.A.L.M.** 1993. *Dynamic Energy Budgets in Biological Systems. Theory and applications in ecotoxicology*. Cambridge University Press.

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

[7] **McCullagh P.** and **Nelder J.A.** 1989. *Generalized Linear Models.* Chapman and Hall, London.

[8] **Morgan B.J.T.** 1996. *Statistics in Toxicology*. Oxford University Press, Oxford.

[9] **OECD** 1998. *Report of the OECD workshop on statistical analysis of aquatic toxicity data.* OECD, Paris.

[10] **Pack S.** 1993. *A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines.* Report to the OECD, Paris.

[11] **Press W.H., Flannery B.P., Teukolsky S.A., and Vetterling W.T.** 1992. *Numerical Recipes in Pascal* Cambridge, Cambridge University Press, Ch. 7.1, p.219

[12] **Seber G.A.F.** and **Wild C.J.** 1989. *Nonlinear Regression.* Wiley, New York.

[13] **Statistical Sciences, S-PLUS** 1995. *Guide to the Statistical and Mathematical Analysis, Version 3,3* Seattle: StatSci, a division of Mathsoft, Inc.

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

[15] **Van Straalen, N.M.** 1997. How to measure no effect. PartII: Threshold effects in Ecotoxicology *Environmetrics*, **8**, 249-253.

[16] **Williams D.A.** 1986. Interval estimation of the median lethal dose. *Biometrics*, **42**, 641-5.