KM 29(1988)
pag 63 - 83



# ESTIMATORS FOR THE LC50

J.A.Hoekstra*

ABSTRACT

In this paper methods to estimate the median lethal concentration (LC50) are reviewed. The performance of the methods is examined by drawing together information from the statistical literature and own simulation studies. Estimation from datasets with limited occurrence of partial response (>0%, <100%) receives particular attention.

Amongst the methods compared are maximum likelihood estimation based on the logit model, the method of Spearman-Kärber and the moving average procedure. The most important difference between the methods lies in the different demands they make on the data to be applicable. In circumstances where they can all be applied results are similar.

*Center for Mathematical Modelling and Statistical Advice, National Institute of Public Health and Environmental Protection, P.O.Box 1, 3720 BA Bilthoven, The Netherlands. Phone: 030-742914

## 1. Introduction

A common measure of the toxicity of a substance to a species is the median lethal (effective) concentration (dose) denoted by LC50 (LD50, ED50). This is the concentration at which 50% of the population is killed within a fixed time. The LC50 is estimated from the results of an acute bioassay. In table 1 an example is given of a bioassay to determine the toxicity of trichlorophenol to the earthworm Eisenia andrei. Relevant for the estimation procedure is the limited occurrence of concentrations with partial mortality (> 0% and < 100%). This can be caused by small variation in tolerance between individuals of an assay and/or lack-of-knowledge of the toxicity of the substance. A second aspect of the data is the incidental occurrence of mortality in the control group, indicating mortality factors other than the toxic substance on test. The dataset is characteristic for many toxicity experiments with earthworms and indeed with many other species as well. Some statisticians would argue that this type of data is too poor for statistical analysis. However, given that it was not known beforehand between which concentrations the LC50 would fall, the data contain a lot of information and it is worthwile to summarize this information succinctly. Of course, sometimes a design providing more information would have been possible with the same resources and within reasonable time. This subject will not be pursued in this paper.

Much is published on methods to estimate the LC50 (e.g. Finney, 1971, 1985; Hamilton et al., 1977; Hoekstra, 1987; James et al., 1984; Kooijman, 1981, 1983; Miller and Halpern, 1980; Racine et al., 1986; Stephan, 1977; Thompson, 1947; Williams, 1986). In this paper the methods will be discussed and relations between methods indicated. I will not attempt to give a complete overview, but the methods known by me to be advocated for routine data-analysis of non-sequential designs will be included. For sequential methods, which can reduce the number of animals to be tested, see Govindarajuluh (1988). The problem of possible dependence between responses will not be discussed (see e.g. Williams, 1982; Moore, 1987). Analysis of multifactor bioassays (Finney, 1971) and modelling the effect of different exposure-times (Ten Berge et al., 1986) is not considered either.

Section 2 will deal with methods based on tolerance distributions, section 3 with tolerance distribution-free methods. Comparisons will be made in section 4, with attention for their usefulness for data as described above. Section 5 contains recommendations.

Table 1 Toxicity of 2,4,5-trichlorophenol to the earthworm Eisenia andrei

| concentration mg/kg | 0 | 0 | 0 | 10 | 10 | 10 | 18 | 18 | 18 | 32 | 32 | 32 | 56 | 56 | 56 | 100 | 100 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| number tested | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| number killed | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 7 | 10 | 10 | 10 | 10 |

Source: C.A.M. van Gestel (1987, unpublished data).

## 2. Methods based on tolerance distributions

### 2.1. The probit and the logit model

P(x), the probability of response in relation to the log concentration x of the toxic substance is most frequently modelled by a probit or a logit model (Finney, 1971; Ashton 1972), formulated as:

$$P(x) = \int_{-\infty}^{\alpha+\beta x} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)\, du$$

and $$P(x) = \frac{1}{1 + \exp(-\alpha - \beta x)}$$, respectively.

Here, $\alpha$ is a location and $\beta$ a scale parameter.
The models can be conceived as tolerance distributions:
each individual is characterized by a log concentration just sufficient to kill it (its "tolerance"). The probability that an individual, randomly drawn from the population, will die at a given log concentration x is:

$$P(x) = \int_{-\infty}^{x} f(u)\, du,$$

in which f(u) is the density function of the tolerance in the population. The probit and logit model assume a normal and logistic density, respectively. Both functions are symmetric. Some asymmetric and more heavy tailed models are discussed by Goedhart (1986). Hamilton (1979) and James et al. (1984) use symmetric, heavy tailed distributions in Monte Carlo studies to compare the robustness of different estimators. Generally, these models need large sample

sizes for adequate estimation, and are seldomly applied to analyse data. More widely used is the extension of the probit or logit model with one parameter to allow for mortality in absence of the toxic substance:

$$P(x) = C + (1 - C) \int_{-\infty}^{x} f(u)\, du.$$

C is the background mortality probability. The LC50 is now defined as the point at which $P = 1/2\ (1 + C)$. For interpretation, Hoekstra (1987) distinguishes between background mortality caused by natural factors and background mortality by the experimental circumstances, like spraying or manipulating the animals. In the last case, to allow extrapolation to a natural situation without artificial mortality, it has to be assumed that the artificially killed animals are not the ones also most sensitive to the poison. If the background mortality is the result of natural processes, however, this extrapolation is unnecessary: the LC50 can be interpreted as the concentration lethal to 50% of the population of individuals that would have survived in absence of the poison. No assumptions about independence of mortality causes need be made. The above model is implicitly assumed in Abbott's formula (1925) to "correct" the data when background mortality is observed. However, direct estimation of the model is preferable to Abbott's correction because the last method does not account for uncertainty in the estimate of the background mortality probability C. To ignore the background mortality can lead to bias in the estimate of the LC50 (Hoekstra, 1987).

## 2.2. Estimation

If a probit or logit model is used, maximum likelihood estimation (MLE) of the log (LC50) $= \mu = -\alpha/\beta$ is the prevalent method. The estimates are obtained either numerically (Finney, 1971) or graphically (Litchfield and Wilcoxon, 1949). The last method seems rather outdated. An alternative estimator is minimum $\chi^2$ (Ashton, 1972).

Confidence intervals can be constructed by applying Fiellers theorem (see Finney, 1978) or the delta-method (see, e.g., Hamilton, 1980). The first method uses normal approximations for $\hat{\alpha}$ and $\hat{\beta}$. The second uses normal approximation for $\hat{\alpha}/\hat{\beta}$ itself, resulting in symmetric intervals. Goedhart (1986) found for 2 investigated data sets that the approximations for $\hat{\alpha}$ and $\hat{\beta}$ were the better. Other parametric methods to obtain confidence limits are the parametric bootstrap (Efron, 1985) and Bayesian procedures (Racine et al.,

1986), which however up to now did not prove their superiority to the simpler Fieller method. Recently, Sanathanan et al. (1987) developed a trimmed logit method in analogy with the trimmed Spearman-Kärber method (see Section 3.1).
All methods described above break down if there is just one or no concentration at which partial kill occurs: the MLE of $\beta$ is then infinite. (Computer programs do not always warn against this situation. A convergence criterion can be met before $\hat{\beta}$ causes overflow, in which case faulty estimates will be given.)
Williams (1986) points out that with one partial kill a confidence interval for the LC50 can still be determined by likelihood methods: the interval is given by the set of values not rejected by a likelihood ratio test. The likelihood ratio test statistic is obtained by subtraction of the deviance under the full model (both parameters estimated by MLE) from the deviance of the model under the null hypothesis (fixed value of $\mu$, $\beta$ estimated by MLE). In case of one partial kill the full model has zero deviance.

## 3. Tolerance distribution-free methods

### 3.1. Spearman-Kärber method

The method of Spearman-Kärber (Spearman, 1908; Kärber, 1931; Finney, 1971) obtains an estimate of $\mu$ as a weighted average of the mid-points between successive log concentrations. The weights are the estimated tolerance densities at the midpoints. Let $x_1 < x_2 < \ldots < x_k$ be the log concentrations and $p_1, p_2, \ldots, p_k$ the observed proportions mortality. The number of individuals tested at $x_i$ is denoted by $n_i$. If $p_1 = 0$ and $p_k = 1$, the Spearman-Kärber estimator (SK) can be written as:

$$\hat{\mu} = \sum_{i=1}^{k-1} (p_{i+1} - p_i) \frac{(x_i + x_{i+1})}{2}$$

If $p_1 \neq 0$ or $p_k \neq 1$ a conventional rule is to extend the series of concentrations with (unobserved) next level, at which p is assumed to be 0 or 1, and to apply the above formula to the extended series.
The formula is an approximation of:

$$E(x) = \int_{-\infty}^{\infty} f(u)\, u\, du$$

if the density vanishes outside the range $(x_1, x_k)$. Equating the above estimator of the mean to the median lethal dose, implies that it is assumed that the underlying tolerance distribution is symmetric.

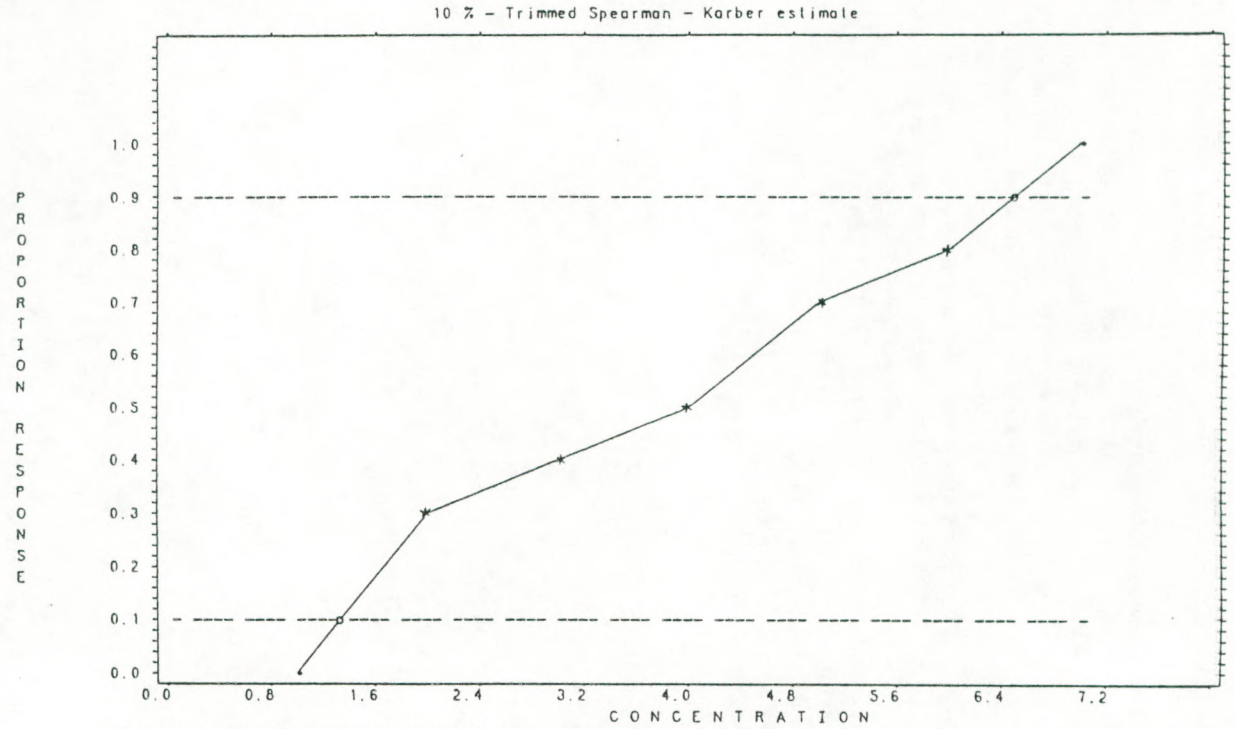


Fig. 1. 10 % – Trimmed Spearman – Karber estimate.
The original SK-estimator is applied to data with proportions between 0.1 and 0.9 (*) supplemented with the estimated values at 0.1 and 0.9 (o). The variance formula is adapted to account for this modification.

The variance of $\hat{\mu}$ is obtained by substituting $p_i(1-p_i)/n_i$ for the variance of $p_i$:

$$\text{Var}(\hat{\mu}) = \sum_{i=2}^{k-1} p_i (1-p_i) (x_{i+1} - x_{i-1})^2/4n_i$$

Sometimes $n_i$ is replaced by $n_i-1$. Two further modifications on the method need to be mentioned. The first is initial monotonization of the proportions by combining the number of responses and subjects of adjacent concentrations whenever the ordening $p_1 < p_2 <....< p_k$ is violated. The monotonized proportions are the MLE's of the true mortality probabilities when no specific tolerance distribution is assumed (Barlow et al., 1972).

The second modification is due to Hamilton (Hamilton et al., 1977, 1978; Hamilton, 1979, 1980). He proposes symmetrically trimmed versions of the Spearman-Kärber estimate in analogy of the trimmed means for continuous data with heavy tailed distributions. The method is exemplified in Figure 1 for the 10% trimmed Spearman-Kärber estimate (SK10%).

Note that the SK50% is equivalent to the median of the empirical tolerance distribution, and amounts to linear interpolation between the two dose-response points that enclose the 50% mortality (Hamilton, 1979). The variance is obtained by application of the delta-method (Hamilton 1979, 1980). In all versions of the Spearman-Kärber method 95% confidence limits are derived as $\hat{\mu} \pm 2$ s.e. $(\hat{\mu})$.

If $p_1 \neq 0$ or $p_k \neq 1$, one can avoid introducing unobserved concentrations with assumed 0% or 100% mortality by the procedure of symmetrical trimming.

## 3.2. Moving average

This method (Thompson, 1947) is seldomly mentioned by statisticians, but popular amongst ecotoxicologists (Stephan, 1977).

It amounts to calculating moving averages of the $p_i$ and $x_i$, followed by linear interpolation between the two averages $x_j^*$ and $x_{j+1}^*$ for which the $p^*$'s enclose 0.50:

$$\hat{\mu} = x_j^* + a.(x_{j+1}^* - x_j^*)$$

$$a = (0.5 - p_j^*) / (p_{j+1}^* - p_j^*)$$

The variance is approximated by the delta-method, and a confidence interval obtained by $\hat{\mu} \pm 2$ s.e. $(\hat{\mu})$. Bennett (1952, 1963) improved the efficiency of

the estimator by weighted averaging according to number of tested individuals and by angular transformation of the proportions.
The span of the averaging procedure should be chosen such that $p^*$'s exist that enclose 0.50. A large span implies the assumption of symmetry.
When "averaging" takes place over span 1, the method is equivalent to SK50%. Averaging over infinite span would equal the untrimmed SK (Finney, 1953).

### 3.3. Other distribution-free methods

Robust estimators for the location of a continuous distribution have been adapted for quantal bioassay data. Table 2 provides a list with references. All estimators are for the centre of symmetry, giving less weight to observations at the tails. Asymptotic considerations and small sample simulations suggest that these estimators are generally comparable with trimmed or untrimmed Spearman-Kärber estimators (Miller and Halpern, 1980; Hamilton, 1979; James et al., 1984). The logistic scores estimator developed by James et al.(1984) seems to be an exception.
It is that value of $\mu$ that satisfies:

$$\frac{1}{n} \sum_{i=1}^{n} J \left\{ \frac{n \left( \hat{F}(x_i) + 1 - \hat{F}(2\hat{\mu} - x_i) \right)}{2n+1} \right\} = 0$$

J is the logit function $J(t) = \log(t/(1-t))$, and $\hat{F}$ is the empirical (monotonized) distribution function. For the logical development of this estimator see James et al. (1984). As far as I know, no estimator of the variance has been developed so that the method is not yet ready for practical use. Shuster and Yang (1975) present the MLE of the smallest $x_i$ such that the response probability is no less than a required level, based on the minimal assumption that the tolerance distribution exists (i.e. p is a non-decreasing function of x). Recently, Glasbey (1987) derived estimators under different sets of assumptions, namely that the tolerance distribution exists, is unimodal, is symmetric, or is symmetric and unimodal. Because no intrapolation between observations is performed, the assumption of symmetry generally does not result in a point estimate, unlike the method of Spearman-Kärber.
Stephan (1977) suggested that in absence of partial kills, the highest concentration with complete survival and the adjacent lowest concentration with complete mortality can be used as limits of an interval with confidence level:

$$(1 - 2\,(1/2)^n) \times 100\%.$$

Here, it is implicitly assumed that all 0% responses precede the 100% responses. This procedure is informally justified and extended in the Appendix.

Table 2

Robust estimators of the LC50

| Name | Reference |
|---|---|
| Trimmed Spearman-Kärber | Hamilton et al., 1977 |
| One-Step Huber M-estimator | Andrews et al., 1972 |
| Sine Curve M-estimator | Andrews et al., 1972 |
| Tukey Biweight | Miller and Halpern, 1980 |
| Logistic Score | James et al., 1984 |
| Hodges-Lehmann | James et al., 1984 |

## 4. Comparison

### 4.1. Introduction

Comparison of LC50-estimators is complicated by the fact that their performance depends quite heavily on the design of the bioassay in relation to the underlying tolerance distribution. A further complication is the fact that most estimators or their variances cannot be calculated for particular datasets. How does one compare the performance of methods if some are not applicable to some realisations of a model/design combination? How does one judge the true confidence level of a method if it does not provide confidence intervals in a number of cases? I will confine the comparison to the maximum likelihood and minimum $\chi^2$ estimators based on the logit model (MLE and MCS), untrimmed and trimmed versions of the Spearman-Kärber estimator (SK), the method of moving average (MA), the logistic scores estimator (LS) and interval estimation by the likelihood ratio method (LR).

In section 4.2 I will summarize articles in which the methods are compared, necessarily with a quite detailed description of the circumstances under which the comparisons are made. In section 4.3 I will add some results of own simulation studies. In section 4.4 conclusions will be drawn.

## 4.2. Literature

The merits of MLE and minimum $\chi^2$ have been disputed extensively (Berkson, 1980). Which method is best seems to depend very critically on the design and the chosen criterion for performance. Generally, minimum $\chi^2$ seems better in terms of estimating $\alpha$ and $\beta$ from the logit model (Berkson, 1955; Smith et al., 1984), however MLE may outperform minimum $\chi^2$ substantially in estimating the log LC50 = - $\alpha/\beta$ (Hamilton, 1979; James et al., 1984).

Finney (1950, 1953) compared the performance of MLE, SK and MA with respect to asymptotic variance and bias under the probit (1950) and logit (1953) model. The experimental design was chosen to be an unlimited series of equally spaced log concentrations each with n observations. In practice, the results can be expected to hold as long as the range of concentrations covers the tolerance density to a large extent. In this situation, the observed response rates should range from approximately 0% to 100%.

Included in the study is the effect of the spacing between designpoints, measured relative to the variance of the tolerance distribution (the variance is inversely related to the slope $\beta$). The asymptotics refer to $n \to \infty$. Also included in the comparison is the effect of the distance of $\mu$ to the nearest designpoint, i.e. the effect of asymmetry of the design with respect to the model.

Important conclusions are:

- SK is identical to logit-MLE (except for discontinuity in the distribution of estimates).
- MA can best be applied with the largest possible span, once a data set has been obtained. It then becomes similar to SK.
- If the design is asymmetrical about $\mu$, SK and MA have a bias independent of n. The wider the spacing of the designpoints, the larger the bias becomes.
- The variance of all three estimators is strongly dependent on the distance from $\mu$ to the nearest designpoint. This instability of the variance increases with the spacing between designpoints. MA with span 1 (using observations at just 2 designpoints) is the most sensitive.

Stephan (1977) discusses the same methods, indirectly referring to the design by mentioning different types of outcomes. He pays attention to the possibility of datasets with no or just one concentration with partial kill. This implies wide spacing of the log concentrations on test relative to the variance of the tolerance distribution. Stephan (1977) advocates the moving average method because of its calculability in cases when $p_1 \neq 0$ or $p_k \neq 1$ (unlike the untrimmed SK which would need undesirable fabrication of data) and when there is just one level with partial kill (unlike MLE). However, he does

not consider the efficiency of MA. This can be low if one is forced by the outcomes to take averages of small span to obtain averages that enclose 50% mortality. This occurs when the observed concentration range covers the tolerance density only partially.

Hamilton (1979) and James et al. (1984) carried out large simulation studies to compare a number of estimators under a variety of symmetric tolerance distribution models. The estimators include the logit MLE and (monotonized) Spearman-Kärber estimations without and with trimming (trimming percentages 5, 10, 20, 50 %). Hamilton uses dosages x = 1,2 ...10, with n = 5, 10, 20. The true value of $\mu$ is fixed at x = 5.5.

The simulated tolerance distribution models include heavy tailed and contaminated distributions. For comparison of both studies I will use the logistic distribution with 10% contamination with another logistic distribution having 100 times the variance of the first. All distributions are scaled so that P (3) = 1- P (8) $\approx$ 0.01. For heavy tailed and contaminated distributions this results in steep tolerance curves. For the above mentioned contaminated logistic distribution no designpoint has a response probability between 0.1 and 0.9. Many datasets with one or no partial kills are the result. Robustness of an estimator under different tolerance distributions is therefore confounded with applicability of the method to datasets with few partial kills. MLE is the only method included in the study that needs at least 2 partial kills. James et al. (1984) used x = 1,2...11, with n = 10, 20 and $\mu$ = 6. The distributions were scaled so that P (3) = 1-P (9) $\approx$ 0.01. However as their design includes $\mu$ as designpoint, appreciable stochastic variation is left at least at this point, even for heavy tailed distributions. The outcomes of the two studies differ remarkably.The efficiencies of the trimmed SK estimators in the study of James et al. (1984) are much lower and not as strikingly different as in Hamilton's article, and sometimes the order is reversed. For example, Hamilton's SK50% has a mean squared error (MSE) of 0.0014 for the contaminated logistic (n=20) and is in this respect 5.6 times better than SK. The tables of James et al. (1984) imply an MSE of 0.033 for SK50%, 0.578 times worse than SK. Smaller but still appreciable differences exist for trimmed Spearman-Kärber estimators with smaller trimming percentages. These differences are attributable to the different position of the dose levels relative to $\mu$ in the two studies. It may be recalled from Finneys study that the variance depends quite heavily on the location of design with respect to $\mu$. This explanation was confirmed by a small simulation study (100 repetitions, n=10) using Hamilton's design and a logit tolerance density with $\beta$ = 6. If $\mu$ was fixed at 5.5 the MSE of SK50% was 0.0035, if $\mu$

was randomly drawn from a uniform distribution between 5 and 6 (100 independent drawings), the MSE increased to 0.0251.
Another consequence of considering just one symmetric position of $\mu$ is that the bias cannot be properly studied. Both studies explicitly mention the probable absence of bias due to the symmetry of the design with respect to $\mu$.

The conclusions of the two simulation studies are as follows.
Hamilton (1979) recommends the 10% or 20% trimmed SK for n = 10,20 because of their favourable MSE under a variety of tolerance models (but we have seen that his MSE's are misleadingly small). James et al. (1984) find that SK5% performs quite well. Heavier trimmed SK's were never more efficient than the untrimmed SK. The logistic scores estimator first presented in this study is overall the best, its efficiency is up to 30% higher than SK under heavy tailed models and less than 20% lower under the logit model.
Miller and Halpern (1980) study asymptotic properties ($n \to \infty$, difference between designpoints $\to 0$ and infinite number of designpoints) of several robust estimators and find that SK10% performs better than SK5% for two very heavy tailed distributions but less for moderately contaminated distributions. Hamilton (1980), using the same design and tolerance distribution as he did before, compared 95% confidence intervals obtained from logit-MLE (delta-method) and trimmed SK's in a Monte Carlo study (n = 5,10,20). The empirical coverages never fell below 91% for any of the methods. I consider this to be close enough to the nominal value. With strongly contaminated steep tolerance distributions MLE, SK, SK5% and SK10% all resulted in conservative confidence intervals, MLE giving the widest intervals in cases where all methods could be applied (>1 partial kill). We expect these conclusions to be valid also if the centre of the design is chosen randomly with respect to $\mu$.
Williams (1986) comes to the conclusion that for small sample sizes MLE-intervals obtained by the Fieller method are conservative even under the logit model itself. He uses designs with a total of 20 to 30 observations spread over 4 to 6 dose levels and different values of the design centre relative to $\mu$. For some simulations the empirical confidence probability of MLE exceeds 98%. The LR-intervals were reported to be liberal, but rarely fell below 93% confidence probability.
The simulation study included models leading to substantial percentages of datasets with one or no concentrations with partial kill.
In summary, the literature seems to indicate for designs that cover most of the tolerance densities and small and moderate sample sizes (10 levels, n = 5, 10) that:

(i) MLE, SK and SK5% do not differ much in terms of MSE. SK5% is perhaps the best because it performs well under a variety of distributions. Only for very heavy tailed tolerance distributions SK10% may outperform the earlier mentioned estimators.

(ii) Empirical confidence levels from MLE, SK5% and LR are reasonably close to the nominal value. For steep or heavy tailed distributions MLE gives the widest confidence intervals.The MLE-intervals obtained with Fiellers method are the most conservative.

(iii)The logistic scores estimator deserves further investigation (performance of variance estimator, confidence intervals etc.)

To confirm some of the above conclusions under the situation that the centre of the design is randomly located with respect to $\mu$, and to include the moving average method (MA) in the comparison, a small simulation study was performed.

4.3 Additional simulation study

MLE, LR, SK and MA with span 8 are compared using as design $x = 1,2,\ldots,10$ with $n = 10$ at each level. The position of $\mu$ is drawn from a uniform distribution between 5 and 6, for each of the 1000 simulated datasets independently. This corresponds to the (still optimistic!) practical situation of knowing that $\mu$ falls in the central interval, but not its exact location. The underlying tolerance distribution was the logistic distribution with $\beta = 1,2,3$ and 6, corresponding to an increasingly wider spacing of the design with respect to the variance. The average percentages response at the 10 designpoints are given in Table 3.

The results are presented in Table 4.

Table 3

Average percentage response in simulation study

| | $\beta$ | | | |
|---|---|---|---|---|
| x | 1 | 2 | 3 | 6 |
| 1 | 1.0 | 0.0 | 0.0 | 0.0 |
| 2 | 3.0 | 0.1 | 0.0 | 0.0 |
| 3 | 7.8 | 0.7 | 0.1 | 0.0 |
| 4 | 18.5 | 5.6 | 1.6 | 0.0 |
| 5 | 36.6 | 28.8 | 21.5 | 10.7 |
| 6 | 62.0 | 71.7 | 78.0 | 87.7 |
| 7 | 80.5 | 94.6 | 98.6 | 100.0 |
| 8 | 91.9 | 99.3 | 100.0 | 100.0 |
| 9 | 97.0 | 99.9 | 100.0 | 100.0 |
| 10 | 99.1 | 100.0 | 100.0 | 100.0 |

Simulation and calculations were thoroughly checked, amongst others by running the programs modified in such a way to obtain results comparable with those of Hamilton (1979) and Williams (1986). The outcomes then agreed with the outcomes reported by these authors. For each method, all statistics are given with reference to those datasets on which the method could be applied. Hence, the results of particularly MLE refer to just a small subset of all datasets for the larger $\beta$'s. The results are scaled with respect to the variance of the tolerance curve.

Table 4

Results of simulation study

| | $\beta$ 1 | 2 | 3 | 6 |
|---|---|---|---|---|
| Datasets without partial response | 0% | 0% | 0.8% | 14.4% |
| Datasets with 1 partial response | 0% | 3.5% | 28.4% | 74.6% |
| Spearman-Kärber | | | | |
| Bias * $\beta$ | -0.0122 | 0.0124 | 0.0123 | 0.0084 |
| MSE * $\beta^2$ | 0.0952 | 0.2028 | 0.2916 | 0.6768 |
| C.I. includes $\mu$ | 95.6% | 94% | 94% | 91.5% |
| Average width C.I. * $\beta$ | 1.22 | 1.74 | 2.13 | 3.24 |
| Moving Average | | | | |
| Bias * $\beta$ | -0.0137 | 0.0122 | 0.0123 | 0.0084 |
| MSE * $\beta^2$ | 0.0989 | 0.2028 | 0.2916 | 0.6768 |
| C.I. includes $\mu$ | 95.5% | 94% | 94% | 91.5% |
| Average width C.I. * $\beta$ | 1.25 | 1.74 | 2.13 | 3.24 |
| Maximum likelihood | | | | |
| Bias * $\beta$ | -0.0127 | 0.0109 | 0.0097 | 0.0441 |
| MSE * $\beta^2$ | 0.0983 | 0.2029 | 0.2969 | 0.5760 |
| F.C.I. incl.$\mu$ | 96.7% | 96.5% | 98.9% | 100% |
| D.C.I. incl.$\mu$ | 95.3% | 93.4% | 95.6% | 98.2% |
| Average width of F.C.I. * $\beta$ | 1.32 | 1.97 | 2.70 | 5.13 |
| Average width of D.C.I. * $\beta$ | 1.23 | 1.70 | 2.21 | 4.04 |
| Likelihood Ratio method | | | | |
| C.I. includes $\mu$ | 95.6% | 94% | 94% | 97% |

C.I. : 95% confidence interval

F.C.I.: 95% confidence interval Fieller method

D.C.I.: 95% confidence interval Delta method

No difference of practical relevance exist between SK and MA8.
This agrees with Finney's (1953) conclusion about the similarity of the 2 methods in this type of situation. Bias played a negligible role in all methods. The empirical confidence probabilities are to my opinion all acceptably close to the nominal value. With a steep logistic curve or equivalently a wide spacing of the concentrations, the methods SK and MA8 give a slightly liberal confidence interval. Hamilton (1980) found the same with a smaller $\beta$ ($\beta$ = 1.838) and n = 5. MLE is not generally applicable in this situation, and is conservative even for the subset of data to which it can be applied. Remarkably, Fiellers method seems no improvement over the simpler delta-method.
With $\beta$ = 6 the LR-interval was conservative, a result that cannot be attributed to sampling error. This contrasts with Williams (1986) findings with smaller sampling sizes. The average width of the intervals was not calculated because of its numerical exactingness (it was only calculated whether the true $\mu$ was not rejected by LR). However, the LR-intervals were not larger than the SK-intervals: with $\beta$ = 6, about 50% of the SK-limits were not rejected by the LR-test.

## 4.4 Conclusions

Lay-out of the designpoints, number of observations and underlying model all influence the performance of the estimators in a rather unpredictable way. Fortunately, the differences between the methods are small as long as the designs covers the tolerance density to a large extent, as was the case in the comparisons discussed in sections 4.2 and 4.3. The most important difference between the methods lies in the different demands they make on the data to be applicable.
The results are summarized as follows:

*Spearman-Kärber method*
Under the logit model, the method is asymptotically equivalent to logit-MLE and hence fully efficient. With small sample sizes (n = 5, Hamilton, 1980), the confidence interval becomes liberal. Large values of $\beta$ have the same effect (section 4.3). The assumption of the conventional Spearman-Kärber estimation namely that responses range from 0% to 100% can be circumvented by using a trimmed version of the estimator. A moderately trimmed version (5% say) is more efficient if the tolerance density function is heavy tailed.

*Logit-MLE*

Also with small samples sizes SK and MLE are comparable, if spacing of the concentrations is fine enough to observe several partial kills. If the spacing is wide MLE becomes inefficient.Confidence intervals based on Fiellers method are more conservative than intervals based on the delta-method.

*Moving average*

This method is most efficient if a large span is used in which case the estimates are practically equal to the Spearman-Kärber estimates. The assumption of symmetry of the tolerance curve implied by SK, logit- and probit MLE, can be relaxed by limiting the span of the averaging procedure.

*Likelihood Ratio confidence interval*

With small sample sizes the empirical confidence probability is less than its nominal value, as was the case with SK (Williams, 1980). Large values of $\beta$ seem to make intervals conservative, at least with n = 10 (section 4.3). It is questionable whether the gain in accuracy of the interval as compared to SK, for example, is worth the extra calculational effort. If so, the confidence limits obtained from SK could be used as starting values for iterative calculation of the likelihood based confidence limits.

*Bias: negligible in all methods*

As long as the underlying tolerance distribution is symmetric, bias plays a negligible role in the performance of all estimators discussed. Symmetry can be enhanced by using an appropriate transformation of the concentrations. Mostly, log transformation is chosen, as was done in this paper, but sometimes there may be reasons to prefer another or no transformation.

## 5. Some practical guidelines

- If the data indicate that only part of tolerance curve is covered by the design (response > 10% at lowest concentration or < 90% at highest concentration) use MLE based on a well fitting model. However do not use MLE if there are less than 3 concentrations with partial response. Use the moving average method with appropriate span for this case.
- If the data indicate asymmetry of the response curve, use MLE based on a well fitting model or the method of moving average with small span (intrapolation between a few central concentrations). Asymmetry can only be observed in large datasets with many partial kills.

- If responses range from 0% to 100% and there are no indications of asymmetry, either of the methods SK, MA with maximal span, or MLE based on a symmetric model (subject to the above restriction) can be used.
- Incidental deaths at low concentrations can indicate background mortality or individuals with high sensitivity to the poison. One of these two possibilities has to be chosen to enable estimation of the LC50. In the first case an Abbott-like modification of the chosen method is appropriate. The second case can lead to choice of an asymmetric response model or - if corresponding incidental survivors at high concentrations are feasible as well - use of a trimmed Spearman-Kärber estimate. In practice it will be very difficult to discriminate between these cases.
- A 95% confidence interval can be constructed by taking the estimate $\pm$ 2 times its standard error. Fiellers method is not recommended, unless the LC50 appears on the edge of the range of tested concentrations.
- Apply a binomial test if no partial kills are observed to obtain a confidence interval (Appendix).

Appendix

We have concentrations $x_1 < x_2 \ldots < x_k$, with n individuals tested at each concentration. The random variable $R_i$ denotes the response at $x_i$, with $r_i$ as its realization. The variables $R_i$, i=1...k, follow a binomial distribution with parameters $P_i$ and n. We assume a non-decreasing response function, i.e. $P_1 \leq P_2 \ldots \leq P_k$.

Suppose, a response percentage above 50% is observed at a certain concentration $x_h$, $h \leq k$. The nulhypothesis $H_0$: $P_h = 0.5$ is tested against the alternative $H_1$: $P_h > 0.5$ with significance level 1/2 $\alpha$. With a non-decreasing response function, all responses at concentrations equal or less than $x_h$ that exceed n/2 plead in favour of $H_1$. Denote the subset of indices out of (1,2...,h) that refer to these observations in favour of $H_1$ by E.

The probability under $H_0$ of obtaining as much or more evidence against $H_0$ as is observed, is less than or equal to:

$$\prod_{j \in E} \text{Prob}\{ R_j \geq r_j \mid P_j = 0.5 \} = \prod_{j \in E} \sum_{1}^{n} \binom{n}{1} 0.5^n$$

If $H_0$: $P_h = 0.5$ is rejected one can proceed with $H_0$: $P_{h-1} = 0.5$, etc. Similar tests can be carried out at the other extreme where the observed responses are much below 50%. The range from the smallest to the largest concentration for which $H_0$ was not rejected constitutes an interval for the LC50 with confidence

level at least 100 $(1-\alpha)$%. For the special case of absence of partial kills, Stephan (1977) inverted this procedure: the level $\alpha$ was calculated for which only the largest concentration with 0% and the smallest with 100% response would be included in the interval (see Section 3.3.). Denote this interval with $(x_s, x_{s+1})$. If n=6, the confidence level is 96.9%. For smaller values of n, one may take a wider interval $(x_{s-b}, x_{s+1+c})$ with confidence level 100 $(1-0.5^{n(b+1)}-0.5^{n(c+1)})$. The symmetric intervals in the table below are chosen in such a way that the confidence level exceeds 95%.

| number of organisms | interval | confidence level |
|---|---|---|
| 1 | $x_{s-5}, x_{s+6}$ | 96.9 |
| 2 | $x_{s-2}, x_{s+3}$ | 96.9 |
| 3 | $x_{s-1}, x_{s+2}$ | 96.9 |
| 4 | $x_{s-1}, x_{s+2}$ | 99.2 |
| 5 | $x_{s-1}, x_{s+2}$ | 99.8 |

Acknowledgements

I would like to thank J.de Bree and S.H.Heisterkamp for discussion on the contents of the Appendix.

References

Abbott,W.S., 1925. A method of computing the effectiveness of an insecticide. J.Econ.Ent., 18, 265 - 267.

Andrews,D.F., Bickel,P.J., Hampel,F.R., Huber,P.J., Rodgers,W.H. and Tukey,J.W., 1972. Robust estimates of location - Survey and advances. Princeton University Press, Princeton.

Ashton,W.D., 1972. The logit transformation. Hafner Publishing Co, New York.

Barlow,R.E., Bartholomew,D.J., Bremner,J.M. and Brunk,H.D., 1972. Statistical inference under order restrictions. Wiley, New York.

Bennett,B.M., 1952. Estimation of LD50 by moving averages. J.Hyg., 50, 157 - 164.

Bennett,B.M., 1963. Optimum moving averages for the estimation of medium effective dose in bioassay. J.Hyg., 61, 401 - 406.

Berge,W.F.ten, Zwart,A. and Appelman,L.M., 1986. Concentration-time mortality response relationship of irritant and systematically acting vapours and gases. Journal of Hazardous Materials, 13, 301-309.

Berkson,J., 1955. Maximum likelihood and minimum $\chi^2$ estimates of the logistic function. JASA, 50, 130-162.

Berkson,J., 1980. Minimum chi-square, not maximum likelihood! Annals of Statistics, 8, 447-487.

Efron,B., 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika, 72, 45 - 58.

Finney,D.J., 1950. The estimation of the mean of a normal tolerance distribution. Sankhya, 10, 341 - 360.

Finney,D.J., 1953. The estimation of the ED50 for a logistic response curve. Sankhya, 12, 121 - 136.

Finney,D.J., 1971. Probit analysis (3rd edn.). The University Press, Cambridge.

Finney,D.J., 1978. Statistical method in biological assay (3rd edn.) Charles Griffin, London.

Finney,D.J., 1985. The median lethal dose and its estimation. Arch. Toxicol., 56, 215 - 218.

Glasbey,C.A., 1987. Tolerance-distribution-free analyses of quantal dose-response data. Appl.Statist., 36, 251-259

Goedhart,P.W., 1986. Statistical analysis of quantal assay data. ITI - TNO rapport, nr. 86, ITI A 31.

Govindarajuluh,Z., 1988. Statistical techniques in bioassay. Karger, New York.

Hamilton,M.A., 1979. Robust estimates of the ED50. JASA, 74, 344 - 354.

Hamilton,M.A., 1980. Inference about the ED50 using the trimmed Spearman-Kärber procedure - a Monte Carlo investigation. Commun. Statist., B9, 235 - 254.

Hamilton,M.A., Russo,P.C. and Thurston,R.V., 1977. Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. Env.Sci. & Technol., 11, 714 - 719. Correction, 1978, 12, 417.

Hoekstra,J.A., 1987. Acute bioassays with control mortality. Water, Air and Soil Pollution, 35, 311-317.

James,B.R., James,K.L. and Westenberger,H., 1984. An efficient R-estimator for the ED50. JASA, 79, 164 - 173.

Kärber,G., 1931. Beitrag zur kollectieven Behandlung pharmakologischer Reihenversuche. Arch. exp. Path. Pharmak.,162, 480 - 487.

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

Kooijman,S.A.L.M., 1983. Statistical aspects of the determinatian of mortality rates in bioassays. Water Res., 17, 749 - 759.

Litchfield,J.T. and Wilcoxon,F., 1949. A simplified method of evaluating dose-effect experiments. J.Pharmac. exp. Ther., 96, 99 - 113.

Miller,R.G. and Halpern,J.W., 1980. Robust estimator for quantal bioassay. Biometrika, 67, 103 - 110.

Moore,D.F., 1987. Modelling the extraneous variance in the presence of extra-binomial variation. Appl. Statist., 36, 8 - 14.

Racine,A., Grieve,A.P., Flühler,H., and Smith,A.F.M., 1986. Bayesian methods in practice: Experiences in the pharmaceutical industry. Appl. Statist., 35, 93 - 150.

Sanathanan,L.P., Gade,E.T., and Shipkowitz,N.L., 1987. Trimmed logit method for estimating the ED50 in quantal bioassay. Biometrics, 43, 825-832

Shuster,J.J. and Yang,M.C.K., 1975. A distribution-free approach to quantal response assays. Can. J. Stats., C, 3, 57 - 70.

Smith,K.C., Savin,N.E. and Robertson,J.L., 1984. A Monte Carlo comparison of maximum likelihood and minimum chi-square sampling distributions in logit analysis. Biometrics, 40, 471-482.

Spearman,C., 1908. The method of 'right and wrong cases' ('constant stimuli') without Gauss's formulae. Q. Jl. Pharm. Pharmac., 20, 380 - 387.

Stephan,C.E., 1977. Methods for calculating an LC50. In: Aquatic toxicology and hazard evaluation, F.L.Mayer and J.L.Hamelink (Eds.), 65 - 84. American Society for Testing and Materials, Philadelphia.

Thompson,W.R., 1947. Use of moving averages and interpolation to estimate median-effective dose. I. Fundamental formulas, estimation of error and relation to other methods. Bact. Rev., 11, 115 - 145.

Williams,D.A., 1982. Extra-binomial variation in logistic linear models. Appl. Statist., 31, 144 - 148.

Williams,D.A., 1986. Interval estimation of the median lethal dose. Biometrics, 42, 641 - 646.

Ontvangen: 28-03-1988
Geaccepteerd: 21-06-1988