

From: Pate Ian GBAP
 Date: 10 August 2001
 Attachment 1: **Direct use of the likelihood function for ED50 estimation**
 Attachment 2: **Figures**
 Subject: **Revised UDP Panel Comments on Evaluation Guidance Questions**

I still have one problem with the revised guidelines and this is regarding the situation with one intermediate dose.

In the case of one intermediate dose the guidelines state that the intermediate dose is to be used as the MLD. Note example 5 page 28 of the guidance document shows an example from this situation where the calculated MLD is not at the intermediate dose - how was this calculated?

Two examples below can be generated by the test process

175	0/2	0/4
550	1/4	3/4
2000	3/3	1/1
MLD	550	550
CI	381-1710	235-852

Both these data would give a point estimate of the MLD as 550. This is difficult to accept given one has a 25% response and one has a 75% response at 550. These data sets should surely be expected to give different estimates of the MLD.

The reason for this problem is covered in a **poster** I presented to the British Toxicology Society in 1989, which I have attached.

The profile likelihood function for this situation (Figure A.2, page 47 of confidence interval description document) is correctly shown as being well behaved. However, the likelihood function itself is not well behaved. My **poster figure 6** shows a three-dimensional plot of the likelihood function which has a ridge at the intermediate dose stretching to a slope of infinity. In simple terms, this is because a perfect fit can be made to the data using a step function i.e. 0% response below intermediate dose, rising from 0-100% response at the intermediate dose and then showing 100% response above it. Consequently, the maximum likelihood estimate of the MLD based on a profile likelihood will always be at the intermediate dose. The chance of a compound exhibiting this steep a dose response is minuscule in practice. For more realistic slope estimates the maximum profile likelihood will not occur at the intermediate dose (unless the observed response is 50%) but will correctly depend on the response observed.

The guidance can easily be changed to calculate the MLD by limiting the slope to the maximum practical value or by taking the mid-point of the profile likelihood confidence interval.

Direct use of the likelihood function for ED50 estimation.

I. Pate.

ICI PLC, Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire, SK10 4TJ.

1. Introduction

The probit, and more recently the logit, model have been used for many years to relate the probability of a response to a chemical to the dose level administered. Interest usually focuses on the estimation of the dose level that is expected to produce a 50% response, the ED50. For the purposes of this poster the logit model will be examined - in practice when interest is centred around the ED50 there is little difference between logit and probit models, although logit models are simpler to handle mathematically.

The logit model can be written

$$\log \frac{p_i}{1-p_i} = - (ED50 + d_i)$$

where p_i is the proportion responding at dose d_i . The model has two parameters, the ED50 and the slope. In general, both the dose d_i and the ED50 are expressed on a \log_{10} scale.

Table 1 contains a data set typical of those generated from an acute toxicity test, where the response is the death of an animal.

Table 1

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5
2000	3.3	5/5

The data points are plotted in Fig.1 along with a fitted logit curve ($\theta=8, ED50=2.9$). The effect on the fitted curve of varying the parameters individually is shown in Fig.2 and Fig.3. The ED50 is reduced in Fig.2a and increased in Fig.2b whilst keeping the slope constant. The slope is reduced in Fig.3a and increased in Fig.3b whilst keeping the ED50 constant. The ED50 parameter determines the position of the dose response curve relative to the dose axis whilst the slope parameter determines the steepness of the dose response curve.

2. The Likelihood Function.

Figs.1-3 demonstrate that some pairs of parameter values provide a dose response curve which fits the collected data more closely than others. The likelihood function provides a numerical measure of support for pairs of parameter values given the collected data. The likelihood function, $L(ED50, \theta)$, is a function of the two unknown parameters ED50 and θ , and can be expressed

$$L(ED50, \theta | \text{data}) = \prod_{i=1}^k \{ (p_i^{\theta r_i}) (1-p_i)^{\theta (n_i-r_i)} \} \quad i=1, \dots, k$$

where k =no. of dose levels tested and

$$P_i = \frac{\exp(- (ED50+d_i))}{1 + \exp(- (ED50+d_i))}$$

The values of the likelihood for the models in Fig.1-3 are shown on the legend to each plot. The best fitting model is that of Fig.1. The maximum likelihood estimates of θ and ED50 are the values which maximise the likelihood function i.e. the best supported values given the data. Fig.1 shows the maximum likelihood solution for the data of table 1.

3. The Shape of the Likelihood Function.

The shape of the likelihood function is shown graphically in Figs.4-8 over a range of data sets chosen to examine

- (i) the effect of sample size
- (ii) the effect of having one or no intermediate doses (i.e. doses which are not 0 or 100% responses)
- (iii) the effect of 0 and 100% responses on the likelihood function.

As our interest generally lies in the ED50 estimate Figs.4b-8b show the likelihood functions rotated so as to view them from along the ED50 axis.

4. Discussion.

The purpose of this poster is to demonstrate graphically the shape of the likelihood function for the type of data common in acute toxicity tests. These are characterised by small numbers of dose levels, usually 3 or 4, with a few, usually 5, animals tested at each dose level. The data often have only one or no intermediate responses. The likelihood function can be used directly to provide point and interval estimates for the ED50.

Point estimates for the ED50 are given by the value with the greatest likelihood and are simply the classical maximum likelihood estimates. The failure of some maximum likelihood programs for data with less than two intermediate responses can be seen from Figs.6 and 7 to be caused by the indeterminacy of the slope (in both cases maximum likelihood estimates occur at a slope of infinity). Point estimates can be obtained by restricting the slope to be less than some predetermined value. This is not unreasonable biologically as an infinite slope corresponds to a dose response model in the unlikely form of a step function.

It is evident that whilst the slope is often poorly defined the range of plausible ED50 values is often tightly bounded. In addition, for data based on as few as five observations at each dose level the shape of the likelihood function is far from normal. Interval estimates which are based directly on the likelihood function, and hence take into account its shape, can be calculated for all the examples in this poster. The intervals are known as profile likelihood or likelihood ratio intervals and whilst the technical details of their calculation are beyond the scope of this poster the motivation for them is evident in Figs.4b-7b. By looking along the ED50 axis we can determine which values of the ED50 are well-supported for any value of the slope, the construction of the interval then requiring only a definition of how well-supported a value needs to be before it is placed in the "likely" interval. Details of the necessary calculations can be found in Williams (1) and Aitkin et al (2).

5. Conclusion.

Direct examination of the likelihood function can provide both point and interval estimates for the ED50 for data based on small numbers of observations at each dose levels and for data containing less than two intermediate responses. The necessary calculations can be programmed easily in GLIM (3) or FORTRAN and with a little more effort using PROC CATMOD in SAS (4).

6. References.

- (1) Williams DA. Interval Estimation of The Median Lethal Dose. *Biometrics*, 1986; 42: 641-645.
- (2) Aitkin M, Anderson D, Francis B and Hinde J. *Statistical Modelling In GLIM*, pp192-194. Oxford: Clarendon Press, 1989.
- (3) Baker R, and Nelder,J. *The GLIM System, Release 3*. Oxford: distributed by the Numerical Algorithm Group, 1978.
- (4) SAS Institute Inc. *SAS User's Guide : Statistics, Version 5 Edition*. Cary,NC: SAS Institute Inc, 1985.

Fig.4.

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/50
500	2.7	10/50
1000	3.0	30/50
2000	3.3	50/50

ED50 = 2.91 Slope = 8.0

95% Confidence Limit for ED50 = (2.85,2.96)

Fig.4 shows the likelihood function for a hypothetical data set with a large number of observations for each dose. The function is well-defined for both the ED50 and the slope, i.e. for both parameters only a narrow range of plausible values exist.

Fig.5.

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5
2000	3.3	5/5

ED50 = 2.91 Slope = 8.0

95% Confidence Limit for ED50 = (2.65,3.10)

Fig.5 shows the likelihood function for a data set with the same response proportions as Fig.4 but now based on only 5 observations per dose level. The likelihood function is now less well-defined, particularly in terms of the slope. The ED50 is still tightly bound although the reduction in sample size has increased the width of the confidence interval.

Fig.6.

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
2000	3.3	5/5

ED50 = 2.70 Slope =

95% Confidence Limit for ED50 = (2.61,3.13)

Fig.6 shows the likelihood function for a data set with only one intermediate response. The indeterminacy of the slope seen to a small extent in Fig.5 is now grossly exaggerated, the likelihood function having a ridge at the intermediate dose. The maximum likelihood estimate occurs on this ridge at a slope of infinity. However, when we restrict interest to the ED50 the indeterminacy of the slope is not apparent and the confidence is little changed from that of Fig.5.

Fig.7.

Dose Log dose Proportion of deaths

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/5
2000	3.3	5/5

ED50 = (1.7,3.3) Slope =

95% Confidence Limit for ED50 = (1.7,3.3)

Fig.6 shows the likelihood function for a data set with no intermediate response. The ridge in the likelihood present in Fig.6 is now apparent for all doses within the range of the 0 and 100% response. There is no unique maximum for the ED50, all values within this range are equally well-supported (at a slope of infinity).

Fig.8.

Dose	Log ₁₀ dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5

ED50 = 2.93 Slope = 6.03

95% Confidence Limit for ED50 = (2.62,4.49)

This example is chosen to highlight the importance of the 100% response (a similar effect can be shown for the 0% response). Whilst values with 0% or 100% responses have infinite logits (or probits) they have a major impact on the likelihood function. In particular, whereas the intermediate responses define the slope of the dose response the 0% and 100% responses are critical in the definition of the location ie the ED50 estimate. The upper confidence limit for the ED50 is now considerably higher than in Fig.5.