



Estimation of the typical lethal dose in acute toxicity studies

Authors: Martin A. Hamilton

This is a postprint of a book chapter that originally appeared in *Statistics in Toxicology* in 1991.

Hamilton, M.A., "Estimation of the typical lethal dose in acute toxicity studies," *Statistics in Toxicology*, 1991, Chapter 4, pp. 61-88.

Made available through Montana State University's [ScholarWorks](https://scholarworks.montana.edu)
scholarworks.montana.edu

91-001

STATISTICS IN TOXICOLOGY

Edited by

Daniel Krewski

*Environmental Health Directorate
Department of National Health and Welfare
Ottawa, Ontario, Canada*

and

Claire Franklin

*Environmental Health Directorate
Department of National Health and Welfare
Ottawa, Ontario, Canada*

1991

(677 pages)

GORDON AND BREACH SCIENCE PUBLISHERS
New York • Philadelphia • London • Paris • Montreux • Tokyo • Melbourne

Chapter 4

Estimation of the Typical Lethal Dose in Acute Toxicity Studies

MARTIN ALVA HAMILTON

I. INTRODUCTION

A. *The Binary Response Bioassay Experiment*

Consider an acute toxicity bioassay in which each test organism is exposed to a known amount of chemical and then observed to determine if a binary (quantal) toxic response subsequently occurs. For example, a minnow could be placed in a beaker of water containing a known concentration of the toxicant and observed 96 hours later to see if it expired, death within 96 hours being the indicator of acute toxicity often used in aquatic toxicology (Sprague, 1969; APHA *et al.*, 1981).

By submitting a few organisms to each of several different levels of a substance, the bioassay experiment can indicate what amount is just sufficient to cause the average subject to respond. The smaller that amount, the greater the toxicity of the substance. In particular, the acute LD50 value, which is the dose lethal to half the organisms, has often been used to classify the degree of toxicity of chemicals (Matsumura, 1985; Racine *et al.*, 1986).

This chapter presents statistical methods for estimating the "typically toxic" amount of a toxicant. These methods are appropriate when the purpose of the experiment is to determine the comparative acute toxicity of substances or to test for acute toxicity as might be required by a routine monitoring program. The methods are applicable to

many field experiments, conducted under circumstances not permitting more sophisticated assays.

Because page limitations do not permit a detailed discussion of each method, an extensive list of references is provided. Some excellent textbooks, monographs, and survey articles have been written on the design and analysis of binary response bioassay data. Among others, Finney (1971, 1978), Hewlett and Plackett (1979), Hubert (1984), and Tsutakawa (1982) are recommended, especially for the classical estimators, such as the probit and logit methods, that are based on parametric models for the dose-response curve. Although this chapter contains a brief survey of those classical estimators, the primary emphasis is on nonparametric and robust methodology.

B. Terminology and Notation

For purposes of this presentation, the word "response" is generic for any well-defined, binary, acute response. The most commonly measured response is death. Only analyses based on the cumulative number of responses are considered here. The time-to-response for each individual is assumed to be unknown or uninformative.

To be acute, the response must occur within a specified (short) time interval. For example, Doull et al. (1980) and National Research Council (1984) classify a response as acute if it occurs within 24 hours of the initial administration of the chemical. In aquatic toxicology, a response is acute if it occurs within 96 hours for fish and other macroinvertebrates (APHA et al., 1981; Sprague, 1969).

The word "dose" will be generic for the level of exposure. Ideally, this would be the peak concentration of the chemical at the physiologic site of action. That concentration is generally unknown, however, and the recorded dose could represent exposures as diverse as the ambient concentration of a substance, the amount of chemical directly injected into the organism, or the concentration of substance contained in the organism's food. The route of administration is an important determinant of the manner in which dose is recorded. Complicated dosing patterns, such as fluctuating concentrations (Mancini, 1983), will not be discussed in this chapter.

In the bioassay experiment, responses are observed at k distinct doses, which are denoted by x_1, x_2, \dots, x_k , arranged in increasing order. In this context, x is the dose value to be used in the statistical

calculations. It can be expressed on any mathematical scale deemed appropriate; e.g., dose could be concentration or logarithm of concentration. It is common practice to base the analysis on the logarithmic scale values (Doull et al., 1980), in which case, the antilog of the final estimate is taken to convert back to the original scale.

Let n_i subjects be exposed to dose x_i and let R_i be a random variable corresponding to the number that respond, $i = 1, \dots, k$. Let $p_i = R_i/n_i$ denote the proportion of the n_i subject that respond at dose x_i . Assume that the response of a subject in no way influences the response of another subject; i.e., the responses are statistically independent. Then the potential outcome (random variable) R_i is statistically distributed with a binomial distribution having mean $n_i \cdot P(x_i)$ and the mean of p_i is $P(x_i)$, the value of the true dose-response curve at x_i . The probability of observing the response pattern $R_1 = r_1, \dots, R_k = r_k$ is given by the product of binomial probabilities:

$$\text{Prob}(r_1, \dots, r_k) = \prod_{i=1}^k \binom{n_i}{r_i} [P(x_i)]^{r_i} [1 - P(x_i)]^{n_i - r_i} \quad (1)$$

All statistical methods described in this chapter are based on the model of equation (1).

C. The True Dose-Response Curve as a Cumulative Distribution Function

Suppose that the subjects used in the bioassay experiment are a random sample from a large population of potential subjects. The unknown, true response associated with dose x is the proportion, denoted by $P(x)$, of subjects in the underlying population that would respond if all were challenged by dose x . The function $P(x)$ formed by varying x over the range of possible dose levels is the true "dose-response curve". Assume that $P(x)$ is monotone, nondecreasing as x increases. Moreover, assume that $P(x)$ approaches zero as x approaches the smallest possible level of exposure and $P(x)$ approaches one as x becomes arbitrarily large. Dose-response curves are usually S-shaped as illustrated by the two curves shown in Figure 4.1.

Just as molecular pharmacologists use the concept of "receptors" and physicists use the concept of "waves" to motivate the mathematical models that fit their data, some biostatisticians use the concept of "tolerance" to motivate calculations based on binary response

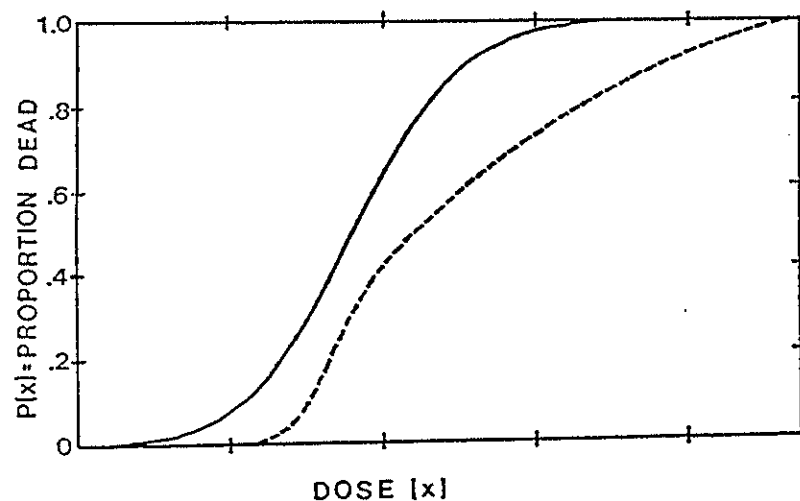


Figure 4.1. Typical shapes for the dose-response function. The solid line is a symmetric dose-response function and the dashed line is a skewed-right dose-response function.

acute toxicity data. Suppose that each subject has a unique "tolerance" dose, determined by the toxicant and the experimental conditions at hand. The tolerance dose is at the threshold for response; that is, at any toxicant dose higher than the tolerance, the subject would respond within the fixed time period, and, at any dose lower than the tolerance, the subject would not respond. Within this conceptual framework, there is a tolerance dose for each subject in the underlying population (Warren, 1971; Goldstein et al., 1974; Finney, 1978). Because a proportion $P(x)$ of the subjects have a tolerance dose less than or equal to dose x , the cumulative distribution function (c.d.f.) of the individual tolerances is exactly the dose-response curve $P(x)$. The concept of tolerance, therefore, allows one to view the "true dose-response curve" as synonymous with the "tolerance distribution function." Much of the discussion in this chapter is motivated by the interpretation of $P(\cdot)$ as a c.d.f.

D. Alternative Definitions of Typically Toxic Dose: The Location of the Tolerance Distribution

The "typically toxic dose" is a parameter chosen to measure the location (central tendency) of the tolerance distribution. For example,

the typically toxic dose for the population of subjects could be represented by the median of $P(\cdot)$. When the response is death, this median lethal dose is denoted LD50. We shall use LD50 here even if the response is something other than death; e.g., a nonlethal pathological sign.

There are a variety of ways, other than the median, to measure the location of the tolerance c.d.f. $P(\cdot)$. The most obvious is the mean. If $P(\cdot)$ is symmetric around the LD50 (i.e., $P(\Theta - x) = 1 - P(\Theta + x)$, for every $x > 0$, where Θ is the LD50), then the mean is identical to the median. On the other hand, if $P(\cdot)$ is skewed (asymmetric), the mean is different from the median. Figure 4.1 shows examples of a symmetric and a skewed-right tolerance distribution. For the latter distribution, the mean is larger than the median.

Although the mean and the median are the most commonly used measures of location, many alternatives exist. One should give some thought to deciding what measure of location is to be estimated. The following mathematical structure provides a common framework for the alternatives.

Let the location measure associated with true tolerance distribution $P(\cdot)$ be denoted by $T(P)$. Given any $P(\cdot)$, T is a rule for finding the numerical value for the location of that $P(\cdot)$. Following Bickel and Lehmann (1975), T is a measure of location if and only if it satisfies three conditions. Consider two toxicants, A and B , and let P_A and P_B be the associated tolerance distributions. The three conditions are:

CONDITION 1. If $P_A(x) \geq P_B(x)$, for every dose x , the location measure must be such that $T(P_A) \leq T(P_B)$.

CONDITION 2. Suppose that, for every subject in the population, dose x of A induces exactly the same response as $ax + b$ dose units of B , where a, b are constants with $a > 0$. Then the location measure must be such that $T(P_B) = aT(P_A) + b$.

CONDITION 3. Condition 2 holds if $a = -1$ and $b = 0$.

Condition 1 is a requirement that, if toxicant A causes at least as much mortality as does toxicant B at every dose level, then the location parameter for A can be no larger than that for B . Condition 2 is the fundamental requirement for toxicity testing applications. It assures that the potency of A relative to B can be calculated from the location parameter values. For example, let toxicant B act as though it were a τ -fold dilution of toxicant A ; i.e., τx units of B cause the same response as x units of A . Then, if T satisfies condition 2, the

potency of A relative to B can be found by $\tau = T(P_B)/T(P_A)$, the ratio of location measures.

An informal survey of bioassay presentations in the biomedical literature reveals that most researchers prefer to express the dose (d) in logarithmic units ($x = \log(d)$) before analysis, in which case condition 2 can still be applied. Suppose $\log(\tau d) (= \log(\tau) + \log(d) = \log(\tau) + x)$ units of toxicant B cause the same response as $\log(d) (= x)$ units of toxicant A . Let P_A and P_B denote the c.d.f.'s for the logarithms of the tolerances. Then, according to condition 2, $\log(\tau) = T(P_B) - T(P_A)$, the difference in location measures (logarithmic dose scale).

As to condition 3, suppose the investigator has recorded the dose as $-x$ and that $P_{-x}(\cdot)$ is the true distribution of tolerances on the negative dose scale. In this case, the true dose-response curve starts at 1.0 and decreases to 0; in fact, the true dose-response curve equals $1 - P_{-x}(\cdot)$. Condition 3 requires that $T(P_{-x}) = -T(P)$.

There are many popular measures of location that satisfy the conditions 1-3; e.g., the median, the mean, and the $100\alpha\%$ trimmed mean for $0 < \alpha < 0.5$. Bickel and Lehmann (1975) have shown that, if P is symmetric around Θ , then Θ is the only parameter satisfying conditions 1-3. Thus a priori knowledge that the tolerance distribution is symmetric allows one to focus on estimators of the LD50, which is the center of symmetry. When P is symmetric, any T that satisfies conditions 1-3 also satisfies $T(P) = \text{LD50}$.

Note that the symmetry of $P(\cdot)$ is dependent on the scale on which dose x is expressed for analysis purposes. If $P(x)$ is symmetric when x is expressed in log units, then $P(x)$ will be asymmetric on the arithmetic unit scale, and conversely. Toxicologists frequently choose a logarithmic scale for x , probably due to empirical evidence that, in many circumstances, $P(x)$ is symmetric when x is expressed in log units. The shape of P depends on the experimental material being studied, however, and the log dose scale will not always produce symmetry. Unless there are compelling reasons to do otherwise, the dose x should be analyzed on the same scale as chosen for recording the data.

II. ESTIMATING THE LOCATION OF $P(\cdot)$

A. Estimating the Median Tolerance

Because the true c.d.f. of tolerances $P(\cdot)$ is unknown, the true median tolerance (LD50) is also unknown. The task is to convert the binary

response bioassay data into an estimate of $P(\cdot)$. Denote the estimate by $\hat{P}(\cdot)$. (Throughout this chapter, a caret $\hat{\cdot}$ over a quantity denotes an estimate of that quantity.) Then estimate the LD50 by that dose at which $\hat{P}(\cdot)$ reaches the ordinate value of 0.5; i.e., the estimate of the LD50 is $\hat{P}^{-1}(0.5)$, the inverse function evaluated at 0.5.

The usual first step in estimating $P(\cdot)$ is to plot each observed response proportion $p_i (= r_i/n_i)$ on the ordinate versus the associated dose x_i on the abscissa. The next step is to fit a smooth, increasing curve through the plotted points. The approaches for fitting a curve through the (x_i, p_i) plot include: (1) a statistical fit based upon a mechanistic model for the dose-response curve $P(\cdot)$, (2) a statistical fit based upon an empirical model, and (3) a model-free data-smoothing technique.

1. The Mechanistic Model Approach. A mechanistic model is a mathematical equation derived from basic biological principles (Warren, 1971; Jacquez, 1972; Goldstein et al., 1974; Hunter and Crowley, 1979; Doull et al., 1980; Altshuler, 1981). The availability of a mechanistic model indicates that the mechanisms by which the substance causes the toxic response are understood; this is rarely the case in acute toxicity bioassays. Thus, although a realistic mechanistic model should be utilized if possible, most experimenters do not have the opportunity to do so. Because any single mechanistic model is of limited applicability to organisms and toxicants other than those for which the model was derived, this approach will not be further discussed here.

2. The Empirical Model Approach, Including Logit and Probit Analyses and the Litchfield-Wilcoxon Method. Underlying the empirical model approach is the assumption that $P(\cdot)$ is closely approximated by a member of a specific parametric family of c.d.f.'s. The most popular models for $P(\cdot)$ are the logistic c.d.f. and the normal c.d.f., which lead to the logit analysis and the probit analysis, respectively (Finney, 1971, 1978).

The logistic model is $P(x) = [1 + \exp\{-(\alpha + \beta x)\}]^{-1}$, which depends on the two unknown parameters α and β . Stated in terms of the logit transform, $\text{logit}\{P(x)\} \equiv \ln\{P(x)/(1 - P(x))\} = \alpha + \beta x$. Let Θ denote the LD50; then $P(\Theta) = 1/2$ and $\text{logit}\{P(\Theta)\} = 0 = \alpha + \beta\Theta$. Thus, $\Theta = -\alpha/\beta$ is the LD50. To estimate Θ , find estimates $\hat{\alpha}$ and $\hat{\beta}$ and then calculate $\hat{\Theta} = -\hat{\alpha}/\hat{\beta}$.

The probit model specifies that $P(\cdot)$ is closely approximated by a normal c.d.f.; viz., $P(x) = \Phi(\alpha - 5 + \beta x)$, where $\Phi(\cdot)$ is the standard

normal c.d.f. and α and β are unknown parameters. Stated in terms of the probit transformation, $\text{probit}\{P(x)\} = \Phi^{-1}(P(x)) + 5 = \alpha + \beta x$. Because $\Phi^{-1}(P(\Theta)) = \Phi^{-1}(1/2) = 0 = \alpha - 5 + \beta\Theta$, the LD50 is $\Theta = (5 - \alpha)/\beta$. To estimate Θ , find estimates $\hat{\alpha}$ and $\hat{\beta}$ and then calculate $\hat{\Theta} = (5 - \hat{\alpha})/\hat{\beta}$.

The number 5 was added to $\Phi^{-1}(P(x))$ for purposes of simplifying hand calculations, a practical consideration when the probit was originally defined decades ago. Now that calculations are performed on computers, it has become popular to drop the 5 from all probit expressions, in which case, the model is sometimes called a normit model (Berkson, 1957).

The logit and probit models for $P(x)$ have been used extensively. For a wide variety of experimental material, toxicologists have observed that, if the dose axis is scaled properly (often a log scale seems to work), then the probit or logit transformed response proportions are linearly related to dose. Without such empirical evidence the use of parametric models may not be justified.

The logistic and normal distributions are symmetric and have very similar shapes. (The solid curve in Figure 4.1 was drawn in the shape of the normal distribution.) The logistic distribution has slightly heavier tails than the normal distribution. The choice between the logit and probit models is largely a matter of convenience; the mathematical difference between the two is so slight as to be of little practical importance (Finney, 1978). Historically, the probit (normal) model has been the most popular of all empirical models. But, perhaps because the logistic model is easier to study analytically, the logit transformation has begun to predominate.

After an empirical parametric model (e.g., a normal or logistic model) is selected, there remains the problem of estimating the unknown parameters α and β in order to estimate the LD50. The two most popular methods of estimation, maximum likelihood and minimum transform chi-squared, can be described as linear regression analyses. Although the dose-response curve $P(\cdot)$ is a nonlinear function of the parameters α and β , it is common practice to create a linear relationship by transforming the response data. For the logistic model, the logit transformation of p_i , $y_i = \ln(p_i/(1 - p_i))$, is an estimate of $\text{logit}\{P(x_i)\}$, which is a linear function of the dose. For the probit model, the probit transformation of p_i , $y_i = \Phi^{-1}(p_i) + 5$, is an estimate of $\text{probit}\{P(x_i)\}$, which is a linear function of dose. For either model, α and β are estimated by the intercept and slope,

respectively, of a weighted linear regression fit to the (x_i, y_i) points. The appropriate weights to use are functions of the unknown $P(\cdot)$, however, and the analysis must be performed with estimated weights.

It turns out that the maximum likelihood estimator of the LD50 for either the logit or the probit model can be calculated by iteratively repeating the weighted linear regression analysis, where the weights used at each step are based on the regression estimate of $P(\cdot)$ at the previous step. The iterative regression analyses continue until the estimate of $P(\cdot)$ is stable (convergent).

The minimum transform chi-squared estimator has been recommended as an alternative to the maximum likelihood estimator (Berkson, 1957). In principle, this estimator can be calculated with a single regression analysis. But it can be improved by smoothing the (x_i, y_i) points with the isotonic regression procedure of Section II.A.3 before computing the regression fit (Hamilton, 1979; James et al., 1984).

The equations for the weighted regression analyses can be found in Finney (1971, 1978, Hewlett and Plackett (1979), Hubert (1984), or McCullagh and Nelder (1983). The calculations of the minimum logit (or probit) chi-squared estimator or of the maximum likelihood estimator are best done on a computer. Widely available computer programs exist to perform these analyses (e.g., Baker and Nelder, 1978; Dixon, 1981 (modules P3R and PLR); SAS Institute Inc., 1982 (procedures FUNCAT and PROBIT)).

There is a formula available for estimating the standard error (SE) of the LD50 estimate associated with the logit (or probit) procedure. Most computer programs automatically calculate the standard error.

The minimum transform chi-squared estimators and the maximum likelihood estimators are asymptotically equivalent. In the small sample size situation, I prefer the maximum likelihood procedure. For either the logit or probit analyses, simulation results indicate that maximum likelihood is preferable to minimum logit (or probit) chi-square for estimating the LD50 (Cramer, 1964; Hamilton, 1977; Smith et al., 1984), although that superiority doesn't necessarily hold for estimating α and β . Cobb and Church (1983) have recently suggested a generalized method-of-moments estimator for α and β of the logistic model that can yield better estimates of the LD50 than maximum likelihood.

An approximation to the probit estimate, known as the Litchfield-Wilcoxon method, is prominent in toxicology (Litchfield and Wilcoxon, 1949; APHA et al., 1981). The Litchfield-Wilcoxon method

reduces computational effort by having the investigator fit the regression lines "by eye", a fact that compromises statistical validity. Although the Litchfield-Wilcoxon method was useful when exact probit calculations were tedious and expensive, *the method is outdated*.

Caution should be exercised in application of analyses based on an empirical model. In particular, the logit or probit method should be used only when the investigator has knowledge that the logistic c.d.f. or the normal c.d.f. adequately approximates $P(\cdot)$. An error in model specification may produce a significantly biased LD50 estimator and/or an invalid standard error calculation. Standard empirical model methods have been shown to be non-robust; i.e., sensitive to a few atypical observations. Robust estimators are discussed in Section II.C.

Methods have been developed for assessing the goodness-of-fit of an empirical model to the observed data and for recognizing outliers (e.g., Brown, 1982; Lemeshow and Hosmer, 1982; Pregibon, 1982; Jennings, 1986; Azzalini et al., 1989; Fowlkes, 1987). Use of these methods, in conjunction with maximum likelihood estimation, should provide a valid LD50 estimate. Whenever the LD50 estimate is based on an empirical model, it is good practice to (i) plot the observed data and estimated dose-response curve and (ii) utilize goodness-of-fit tests and outlier detection methods. Proper application of the latter procedures, however, requires some statistical training and experience. The need for validity checks is a serious disadvantage of the standard empirical model approaches, especially for the routine analysis of many bioassay experiments or for estimation in other situations where a statistician is not available.

In an effort to produce empirical models that are more widely applicable, statisticians have proposed various generalizations of the normal and logistic models (Prentice, 1976; Copenhaver and Mielke, 1977; Aranda-Ordaz, 1981; Guerrero and Johnson, 1982). These more general models can assume a variety of shapes, including asymmetric shapes, but may require that three of four parameters be estimated before the LD50 can be found. Maximum likelihood estimation techniques are usually suggested. Perhaps because associated computer programs are not yet widely available, these new models have not been extensively used. As with the logit and probit models, the main justification for the appropriateness of these generalized models is empirical.

3. Model-free Data-smoothing Techniques. Nonparametric methods for estimating $P(\cdot)$ include isotonic regression (Tsutakawa, 1982), the

moving average (Finney, 1978) and kernel smoothers (Copas, 1983). Consider the estimate of $P(\cdot)$ formed by plotting the k points (x_i, p_i) , $i = 1, \dots, k$, and connecting successive points with straight lines. This is an intuitively appealing model-free estimate. But it possesses two important deficiencies. First, because it is possible for p_i to be larger than p_{i+1} , the estimated dose-response curve may decrease over a range of doses. Second, if the n_i 's are small, the estimated dose-response curve may oscillate wildly. The smoothing techniques presented below attempt to retain the model-free characteristic while producing monotone nondecreasing, smooth estimated dose-response curves.

The isotonic regression approach adjusts the piece-wise linear curve formed by the (x_i, p_i) points so that it is nondecreasing. If $p_i \leq p_{i+1}$, $i = 1, \dots, k - 1$, no adjustment is made. If $p_i > p_{i+1}$, then these two response proportions are replaced by an average proportion; specifically, the (x_i, p_i) and (x_{i+1}, p_{i+1}) points are replaced by (x_i, \bar{p}) and (x_{i+1}, \bar{p}) , respectively, where $\bar{p} = (r_i + r_{i+1})/(n_i + n_{i+1})$. This procedure becomes more complicated if the initial plot decreases at more than one place. Computer algorithms for performing isotonic regression calculations are available (Wright, 1978; Dykstra, 1981; Murray, 1983). Although the median of the curve, smoothed by isotonic regression, is often an inefficient estimator of the LD50, the smoothed curve is the first step in calculating other useful estimators (Church and Cobb, 1973; Hamilton et al., 1977; James et al., 1984).

Equation (2) is a formula for calculating the estimated standard error of the isotonic regression estimated LD50 (Hamilton, 1980). Limited simulation study indicates that (2) is conservative (i.e., the estimated standard error is positively biased), but is otherwise valid if there are partial kills among at least 10 subjects per dose at the dose levels nearest the estimated LD50. Let $L = \max\{i : \bar{p}_i \leq 0.5, i = 1, \dots, k\}$ and let $U = \min\{i : \bar{p}_i \geq 0.5, i = 1, \dots, k\}$; then the estimated standard error of $\hat{\theta} = \hat{P}^{-1}(1/2)$ is the square root of $\text{Var}(\hat{\theta})$, where

$$\begin{aligned} \text{Var}(\hat{\theta}) = & (x_U - x_L)^2 \{ [(\bar{p}_U - 0.5)^2 / (\bar{p}_U - \bar{p}_L)^4] \bar{p}_L (1 - \bar{p}_L) / n_L \\ & + [(0.5 - \bar{p}_L)^2 / (\bar{p}_U - \bar{p}_L)^4] \bar{p}_U (1 - \bar{p}_U) / n_U \}. \quad (2) \end{aligned}$$

Schmoyer (1984) has shown how to adjust the plot of the raw data so that the resulting curve is S-shaped as well as being nondecreasing.

Table 4.1
OBSERVED AND ADJUSTED PROPORTIONS FOR A DIESEL
FUEL AEROSOL EXPERIMENT (SCHMOYER 1984)

Dose (x) h-mg/l	Number (n) rats tested	Number (r) dead	Proportion		
			Observed $p = r/n$	Isotonic adjusted	S-shaped adjusted
8	30	0	0	0	0
16	40	1	0.025	0.025	0.025
24	40	2	0.050	0.050	0.050
28	10	5	0.500	0.425	0.390
32	30	12	0.400	0.425	0.448
48	20	16	0.800	0.733	0.677
64	10	6	0.600	0.733	0.892
72	10	10	1	1	1

His method would be the appropriate extension of isotonic regression if one were sure that $P(\cdot)$ is S-shaped (in a sense defined by Schmoyer). No formulas are available yet for calculating the standard error of the associated LD50 estimate.

Table 4.1 presents an example used by Schmoyer (1984) to illustrate the isotonic and S-shaped adjustments. The data are from an experiment performed by Dalbey and Lock (1982) to assess the lethality of diesel fuel aerosol smoke screens on rats. Note that the observed proportions decrease at two places. The LD50 estimates based on isotonic and S-shaped adjusted proportions are 35.9 ($SE = 3.7$) and 35.6, respectively. Figure 4.2 shows the data of Table 4.1 and the isotonic and S-adjusted dose-response curves.

There are many different moving average procedures, depending on the span chosen and on how the response proportions are transformed before averaging (APHA *et al.*, 1981; Chanter, 1975; Harris, 1959). The most popular moving average uses a span of three and averages the angular transformed proportions, $y_i = \sin^{-1}(\sqrt{p_i})$. Although the median of the moving average smoothed curve is a statistically inefficient estimator of the LD50 (Finney, 1978), some investigators have found that it is informative (Stephan, 1977). Methods for calculating a standard error of the LD50 estimate are given in Bennett (1952). See Engeman *et al.* (1986) for a recent evaluation of the moving average procedure.

Isotonic regression and moving average smoothing are easy methods to understand, easy to calculate, and widely applicable. A combination of isotonic regression and moving averages has been used,

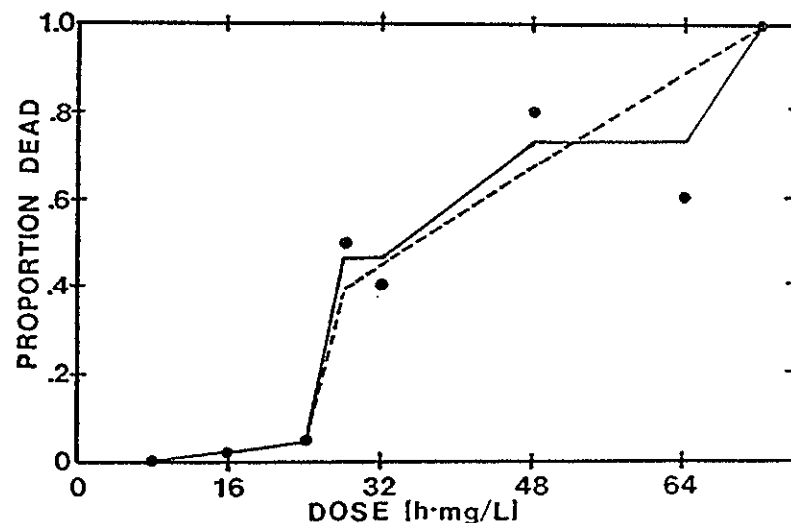


Figure 4.2. A plot of the data shown in Table 4.1. The circles indicate the observed response proportions; the solid line is the isotonic regression smoothed estimate of $P(x)$; and the dashed line is the S-shaped smoothed estimate of $P(x)$ as calculated by Schmoyer (1984).

although in a different context, to smooth quantal response data (Wolfe *et al.*, 1981).

The kernel method is a generalization of the moving average procedure, motivated by kernel techniques for estimating probability density functions. This approach is relatively new, but looks quite promising (Müller and Schmitt, 1988; Staniswalis and Cooper, 1988). One specific kernel estimate of $P(\cdot)$ is

$$\hat{P}(x) = \left[\sum_{j=1}^k r_j e^{-(h/2)(x-x_j)^2} \right] / \left[\sum_{j=1}^k n_j e^{-(h/2)(x-x_j)^2} \right], \quad (3)$$

where h is a smoothness parameter determined subjectively (Copas, 1983) or determined from the data by a cross-validation technique (Kappenman, 1987). Initial work by Kappenman suggests that the median, $\hat{P}^{-1}(0.5)$, of the kernel-smoothed dose-response curve is a robust, reliable estimator of the LD50. To ensure that $\hat{P}(\cdot)$ is non-decreasing, it may be necessary to perform an isotonic regression

adjustment to the data after calculating the kernel-smoothed function. Standard error formulas are not yet available.

B. Estimating the Mean Tolerance

1. *The Spearman-Kärber Estimator.* The leading estimator of the mean tolerance is the Spearman-Kärber (SK) estimator, which can be viewed as the mean of the isotonic regression estimate of $P(\cdot)$ (Brown, 1970; Tsutakawa, 1982). The SK estimator is a nonparametric estimator; one need not choose a family of distributions on which to base the analysis as is required, for example, with the logit and probit methods. If the tolerance distribution is symmetric, then the mean and median are identical and the SK estimator is also an estimate of the LD50. Thus the SK estimator could be used to estimate the LD50 if one believed the dose-response curve could be closely approximated by a logistic or a normal distribution, both of which are symmetric. In fact, the SK estimator has been shown to be a close approximation to the logistic model maximum likelihood estimator and to be nearly as efficient as the probit or logit estimators if $P(\cdot)$ has a normal or logistic shape (Brown, 1961; Church and Cobb, 1973; Miller, 1973; Hamilton, 1979; Miller and Halpern, 1980; Morton, 1981). The validity of SK as an estimator of the mean is in no way affected, however, if the data do not truly follow a normal or logistic model. It is easy to calculate SK and the associated standard error of the estimate (see equations (4) and (5)). For these reasons, I prefer the SK method to logit or probit methods for estimating the typically toxic dose.

The SK estimator and associated standard error can be calculated as shown in equations (4) and (5) if $p_1 = 0$ and $p_k = 1$. When the isotonic adjusted response proportions don't run from 0 to 1, inclusive, the following modifications are standard. If $p_1 > 0$, choose a dose $x_0 (< x_1)$, associate with it the artificial response proportion $p_0 = 0$, and change the lower limit of summation in (4) and (5) to 1. Similarly, if $p_k < 1$, choose a dose $x_{k+1} (> x_k)$, associate with it the artificial response proportion $p_{k+1} = 1.0$, and change the upper limit of summation in (4) to $k + 1$ and the upper limit in (5) to k . The specification of x_0 and x_{k+1} implies that the investigator has (a priori) knowledge that no subjects have tolerances below x_0 or above x_{k+1} . The necessity for an occasional subjective adjustment to equations

(4) and (5) can be considered a deficiency of the method (Hamilton 1979).

$$SK = \sum_{i=2}^k (p_i - p_{i-1})(x_i + x_{i-1})/2 \quad (4)$$

The estimated standard error of SK is the square root of the estimated variance,

$$\text{Vâr}(SK) = \sum_{i=2}^{k-1} (x_{i+1} - x_{i-1})^2 p_i (1 - p_i) / (4n_i). \quad (5)$$

For the data of Table 4.1, the SK estimate is 41.3 ($SE = 2.6$).

2. *The Reed-Muench and Dragstedt-Behrens Methods.* The Reed-Muench and Dragstedt-Behrens nonparametric methods, which are described in Finney (1978), have been used in some acute toxicity tests to estimate the LD50. As shown in Miller (1973), the computations associated with these two methods are quite similar to, but more complicated than, those associated with the SK estimator. Miller demonstrated that the two methods are estimating the mean tolerance which, as has previously been noted, is equal to the LD50 if $P(\cdot)$ is symmetric. Both large sample (Miller, 1973) and small sample (Finney, 1978) evaluations of these estimators have clearly established the superiority of the Spearman-Kärber procedure. Apparently the continued use of the Reed-Muench and/or Dragstedt-Behrens methods in some areas of toxicology is motivated only by tradition.

C. Estimating Robust Measures of Location

1. *Robust Measures: Definition and Rationale.* A robust measure of location has the property that a slight alteration in the tolerance distribution is accompanied by no more than a slight change in the measure. A non-robust measure is greatly influenced by minor perturbations in P . If one does not confine attention to a robust measure of the typically toxic dose, it can happen that one or two aberrant responses can dramatically affect the analysis. In this section, we will confine attention to robust measures of location, $T(P)$, which satisfy a fourth condition in addition to the three conditions listed in Section I.D:

CONDITION 4. If P_A and P_B are nearly equal, then $T(P_A)$ and $T(P_B)$ are nearly equal; i.e., T is continuous.

A rigorous mathematical definition of condition 4 can be found in Huber (1981). Hamilton (1979) discusses robustness in the context of binary response bioassay. The basic idea is that a few unusually resistant animals (or a few unusually sensitive animals) should not greatly affect the measure of location, nor should a few outliers in the data, due possibly to errors in dosing, errors in recording outcomes, etc. In my experience, it is not unusual to see anomalous responses in acute toxicity bioassay data, particularly when a large number of experiments are being conducted on a routine basis.

The median and the trimmed mean are robust measures of location. Not all location measures satisfy condition 4. In particular, the mean is non-robust, as it can be greatly affected by a few anomalous tolerances. This point is illustrated with numerical examples in Section II.C.2.

The robust estimation approach to analyzing acute toxicity test data has three steps: (i) Pick a robust measure of location, T ; (ii) using a model-free smoothing method, estimate $P(\cdot)$; and (iii) estimate location with $T(\hat{P})$. Estimators found in this manner require minimal statistical assumptions (Hamilton, 1979; Miller and Halpern, 1980; James et al., 1984) and, therefore, they are widely applicable.

The logit and probit methods of Section II.A.2 estimate the LD50, the median of P , which is a robust measure of location. These methods, however, do not follow steps (ii) and (iii) for robust estimation; model dependent weighted least squares regression fits are used to estimate P and the LD50. The standard logit and probit estimators are known to be non-robust (Hamilton, 1979; James and James, 1983); numerical examples illustrating this non-robustness are provided below. Sanathanan et al. (1987) have shown how to use trimming to make the logit procedure less sensitive to unusual observations in the tails of the dose-response curve.

2. *Estimating the Trimmed Mean of the Tolerance Distribution: The Trimmed Spearman-Kärber Estimator.* The trimmed mean of the tolerance distribution is a robust measure of location for which the associated estimator is particularly intuitive, informative, easily calculable, and has good precision. The $100\alpha\%$ trimmed mean is the mean of the middle $100(1 - 2\alpha)\%$ of the distribution, $0 < \alpha < 0.5$; the lowest $100\alpha\%$ and the highest $100\alpha\%$ of the distribution are "trimmed". For analysis of bioassay data, the associated estimator is called the $100\alpha\%$ Trimmed Spearman-Kärber ($100\alpha\%$ TSK) esti-

mator (Hamilton et al., 1977), which is defined by equations (6) and (7) for the case where P is estimated by isotonic smoothing. Although other types of smoothing could be employed, the many toxicologists who have used the TSK have followed equations (6) and (7).

To calculate the $100\alpha\%$ TSK estimate, one must first decide on the trimming coefficient α , which is the proportion of extreme values to be trimmed from each tail of the tolerance distribution, $0 < \alpha < 0.5$. In my experience, $\alpha = 0.10$ is a good choice. If one chose $\alpha = 0$, the result would be the SK estimator of the mean tolerance, which is non-robust. The limiting case as α approaches 0.5, is the median of $\hat{P}(\cdot)$, which is robust. If one believes that P is asymmetric, then all analyses to be compared should employ the same trimming coefficient. If P is known to be symmetric, then location estimators are estimating the same parameter (the center of symmetry, LD50) and comparisons among assays are appropriate even if different α 's are used.

Let \bar{p}_i , $i = 1, \dots, k$ be the response proportions smoothed by isotonic regression. Calculate

$$\begin{aligned} L(\alpha) &= \max\{i : \bar{p}_i \leq \alpha, i = 1, \dots, k\}; \\ U(\alpha) &= \min\{i : \bar{p}_i \geq 1 - \alpha, i = 1, \dots, k\}; \end{aligned} \quad (6)$$

$$\hat{P}^{-1}(\alpha) = x_{L(\alpha)} + [x_{L(\alpha)+1} - x_{L(\alpha)}][\alpha - \bar{p}_{L(\alpha)}] / [\bar{p}_{L(\alpha)+1} - \bar{p}_{L(\alpha)}];$$

and

$$\begin{aligned} \hat{P}^{-1}(1 - \alpha) &= \\ & x_{U(\alpha)-1} + [x_{U(\alpha)} - x_{U(\alpha)-1}][1 - \alpha - \bar{p}_{U(\alpha)-1}] / [\bar{p}_{U(\alpha)} - \bar{p}_{U(\alpha)-1}]. \end{aligned}$$

Then

$$\begin{aligned} 100\alpha\% \text{ TSK} &= 0.5\{[\bar{p}_{L(\alpha)+1} - \alpha][\hat{P}^{-1}(\alpha) + x_{L(\alpha)+1}] \\ &+ \sum_{i=L(\alpha)+1}^{U(\alpha)-2} [\bar{p}_{i+1} - \bar{p}_i][x_i + x_{i+1}] \\ &+ [1 - \alpha - \bar{p}_{U(\alpha)-1}] \\ &\cdot [\hat{P}^{-1}(1 - \alpha) + x_{U(\alpha)-1}]\} / \{1 - 2\alpha\}. \end{aligned} \quad (7)$$

This 100 α % TSK estimate is calculable only if $\bar{p}_1 \leq \alpha$ and $\bar{p}_k \geq 1 - \alpha$; if these inequalities are not satisfied, a modification of equation (7) similar to that described for the SK estimator (equation (4)) is required.

The standard error of the 100 α % TSK can be estimated by the square root of the estimated variance given by equations (8) and (9) (Hamilton, 1980). A computer program (EPA 1986) is available for performing TSK calculations on a personal computer.

Denote the $L(\alpha)$ and $U(\alpha)$ of equation (6) simply by L and U . Define

$$\begin{aligned} V_1 &= [(x_{L+1} - x_L)(\bar{p}_{L+1} - \alpha)^2/(\bar{p}_{L+1} - \bar{p}_L)^2] \bar{p}_L(1 - \bar{p}_L)/n_L; \\ V_2 &= [(x_L + x_{L+2}) + (x_{L+1} - x_L)(\alpha - \bar{p}_L)^2/(\bar{p}_{L+1} - \bar{p}_L)^2] \\ &\quad \cdot \bar{p}_{L+1}(1 - \bar{p}_{L+1})/n_{L+1}; \\ V_3 &= \sum_{i=L+2}^{U-2} (x_{i-1} - x_{i+1})^2 \bar{p}_i(1 - \bar{p}_i)/n_i; \\ V_4 &= [(x_{U-2} - x_U) + (x_U - x_{U-1})(\bar{p}_U - 1 + \alpha)^2/(\bar{p}_U - \bar{p}_{U-1})^2] \\ &\quad \cdot \bar{p}_{U-1}(1 - \bar{p}_{U-1})/n_{U-1}; \\ V_5 &= [(x_U - x_{U-1})(1 - \alpha - \bar{p}_{U-1})^2/(\bar{p}_U - \bar{p}_{U-1})^2] \\ &\quad \cdot \bar{p}_U(1 - \bar{p}_U)/n_U; \\ V_6 &= \{[(x_U - x_{L+1})(1 - \alpha - \bar{p}_U)^2/(\bar{p}_U - \bar{p}_{L+1})^2] \\ &\quad - [(x_{L+1} - x_L)(\alpha - \bar{p}_L)^2/(\bar{p}_{L+1} - \bar{p}_L)^2] \\ &\quad + (x_L - x_U)\}^2 \bar{p}_{L+1}(1 - \bar{p}_{L+1})/n_{L+1}. \end{aligned} \quad (8)$$

If $U - L \geq 4$, then

$$\text{Var}(100\alpha\% \text{ TSK}) = (V_1 + V_2 + V_3 + V_4 + V_5)/[2(1 - 2\alpha)]^2;$$

if $U - L = 3$, then

$$\text{Var}(100\alpha\% \text{ TSK}) = (V_1 + V_2 + V_4 + V_5)/[2(1 - 2\alpha)]^2;$$

if $U - L = 2$, then

$$\text{Var}(100\alpha\% \text{ TSK}) = (V_1 + V_5 + V_6)/[2(1 - 2\alpha)]^2;$$

and if $U - L = 1$, then

$$\begin{aligned} \text{Var}(100\alpha\% \text{ TSK}) &= (x_U - x_L)^2 \\ &\quad \cdot \{[(\bar{p}_U - 0.5)^2/(\bar{p}_U - \bar{p}_L)^4] \bar{p}_L(1 - \bar{p}_L)/n_L \\ &\quad + [(0.5 - \bar{p}_L)^2/(\bar{p}_U - \bar{p}_L)^4] \bar{p}_U(1 - \bar{p}_U)/n_U\}. \end{aligned} \quad (9)$$

For the data of Table 4.1, the 10% TSK estimate is 39.9 (SE = 3.1). The plot of the data (Figure 4.2) suggests that the tolerance distribution is skewed right. Note that the 25th percentile is much closer than is the 75th percentile to the 50th percentile. Consequently, the 10% TSK and an LD50 estimate are probably estimating different measures of location. The same choice of estimator should be made for all experiments to be compared; that is, if 10% TSK were used here, then the other experiments should also be summarized by their 10% TSK values.

As an illustration of the effect of an anomalous response on some of the estimators discussed so far, consider the fictitious bioassay results of Table 4.2. The only difference between the two sets of data occurs at dose $x = 3$ where experiment B exposed one more subject than experiment A and that extra subject responded. The positive response of the extra subject could be considered atypical, perhaps due to a data recording error, an error in dose administration, or the fact that the extra subject was unusually sensitive; the true reason probably is not discernable.

Table 4.2 includes the logistic model maximum likelihood estimate (MLE), the SK estimate, the 5% TSK estimate, the 10% TSK estimate, and the median of the isotonic regression dose-response curve (50% TSK). The MLE and SK estimates are 0.10 dose units lower for experiment B than for experiment A, but the robust estimates are about the same for B as for A. For experiment B the MLE and SK estimates are less than 6.0 and the robust estimates are all greater than 6.0. Because the response at $x = 6.0$ is only $p = 0.3$, one might reasonably expect the typically toxic dose to be larger than $x = 6.0$. Clearly, MLE and SK are more greatly affected by the anomalous response.

3. Other Robust Estimators of Location. Other robust estimators are described and evaluated in Hamilton (1979), Miller and Halpern

Table 4.2
RESULTS OF TWO FICTITIOUS BINARY RESPONSE
EXPERIMENTS WITH ASSOCIATED ESTIMATES OF THE
LOCATION OF THE TOLERANCE DISTRIBUTIONS
(CF. HAMILTON 1979)

(a) Results							
Experiment A				Experiment B			
Dose (x)	No. (n) tested	No. (r) respond	Fraction $p = r/n$	Dose (x)	No. (n) tested	No. (r) respond	Fraction $p = r/n$
2	10	0	0	2	10	0	0
3	10	0	0	3	11	1	0.091
4	10	0	0	4	10	0	0
5	10	2	0.200	5	10	2	0.200
6	10	3	0.300	6	10	3	0.300
7	10	10	1	7	10	10	1

(b) Estimates		
Estimator	Experiment	
	A	B
Logistic model maximum likelihood	6.00	5.90
Spearman-Kärber (SK)	6.00	5.90
5% Trimmed Spearman-Kärber (5% TSK)	6.05	6.03
10% Trimmed Spearman-Kärber (10% TSK)	6.10	6.09
50% Trimmed Spearman-Kärber (median isotonic)	6.29	6.29

(1980), James and James (1983), and James *et al.* (1984). The logistic scores estimator of James *et al.* (1984) appears to be particularly reliable, at least if the true dose-response curve is symmetric.

Improvements to these robust estimation procedures may be realized by calculating $\hat{P}(\cdot)$ with some smoothing procedure other than isotonic regression (e.g., equation (3)). One might reasonably expect that an estimator would have greater precision if it were based on a better smoothing procedure. This conjecture remains to be studied.

III. MISCELLANEOUS TOPICS

A. Adjustment for Background Effects

Many bioassay experiments include a control group, consisting of subjects not exposed to the substance. The methods discussed in the previous sections need to be modified in some way if the control response proportion is nonzero. When the dose-response curve is to

be estimated by fitting an empirical model, such as the probit model, the standard approach is to adjust the observed proportions by Abbott's correction (APHA *et al.*, 1981) prior to doing the regression calculations. Hasselblad *et al.* (1980) outline a computational method for accomplishing the same goal as does Abbott's correction, but in a more satisfactory way. Their presentation gives details only for the probit model, but the approach could be adapted to any empirical model.

The proportionate response among the controls is sometimes attributed to "background effects." Before deciding how to correct for these effects, it is important to consider their causes. The Abbott correction and the Hasselblad *et al.* methods assume that background effects, whatever their origin, and those effects due to the toxicant under study combine *independently*. This assumption is not always correct. Consider, for example, a situation where the background effects are partially (or completely) due to small amounts of the toxicant, or of a chemical structurally similar to the toxicant. A subject exposed to the test concentration of the toxicant is in fact challenged with a concentration equal to the sum of the background and test concentrations. In this case, the background effects enter into the model in a very different manner than they do under the independence assumption and Abbott's correction may not be appropriate. See Cornfield *et al.* (1978), Hoel (1980), and their references for a detailed discussion of this issue.

Fortunately, for purposes of acute toxicity tests, different methods of correcting for background effects usually do not greatly alter the estimate of location. Most investigators, therefore, choose to use Abbott's correction (or a similar technique) whenever background effects are observed in an acute toxicity bioassay.

B. Extra-binomial Variation

Equation (1) is based on the assumption that, if n subjects are exposed to dose x , then the number responding (R) has a binomial distribution with mean $nP(x)$. In some circumstances, it has been observed that the variance of R is larger than could be attributed to binomial variation. For example, if two different tanks of n fish each contain concentration x , the responses in the two tanks could be significantly different. One might conclude that a "tank effect" influences the data

and that, although the mean number of responses may be the correct value $nP(x)$, the variance of R is greater than the binomial variance $nP(x)[1 - P(x)]$.

In such circumstances, modifications to the methods of estimation are required in order to quote a relevant standard error. Examples of such modifications of the logistic maximum likelihood procedure are given in Williams (1982) and Pregibon (1982).

Bayesian Approaches

There are a number of practical reasons why one might consider Bayesian approaches to the analysis of acute toxicity test data. The investigator may have prior information about the dose-response curve. For example, the experiment at hand may be the latest in a series of similar monitoring experiments, where the true dose-response curve varies from experiment to experiment. In such cases, one might desire a Bayesian analysis by which the prior information is blended with the data to estimate the dose-response curve.

When the acute toxicity test is performed with few subjects and/or widely spaced doses, classical estimates may not be calculable. Bayesian techniques, however, can be used in such situations (Racine *et al.*, 1986).

The Bayesian approach to bioassay can be viewed as another form of smoothing to estimate the dose-response curve. Even if one does not accept the Bayesian motivation, Bayesian curves can provide useful alternatives to the model-free data-smoothing techniques of Section II.A.3.

Finally, Bayesian reasoning has had a major role in statistical derivations of optimal experimental designs for acute toxicity bioassays. In order to understand the design recommendations, some knowledge of Bayesian bioassay may be required.

A select list of references to Bayesian bioassay procedures is: Ramsey (1972), Stewart (1979), Brunk (1981), Disch (1981), Kuo (1983), Ishiguro and Sakamoto (1983), and Racine *et al.* (1986). For the most part, the estimates described in these papers are plagued by computational difficulties. With rapidly diminishing costs for quality computer time, they probably are now practical. Nevertheless, Bayesian methods in acute toxicity testing remain the exception rather than the rule.

D. Design of Acute Toxicity Bioassays

The design of the bioassay is determined by the number of dose levels (k), the placement of the dose levels (x_1, \dots, x_k), and the number of subjects per dose level (n_1, \dots, n_k). There may be constraints placed on the design; e.g., a maximum total number of subjects and/or a maximum number of dose levels. To determine the optimal design, one must specify the criterion to be applied and the estimator to be used. For example, one might want the design that minimizes the standard error of the Spearman-Kärber estimator.

Design recommendations are often stated in terms of a prior guess of the location of the tolerance distribution $P(\cdot)$. Preliminary experimentation with a small number of animals is often the basis for that prior information. A popular method for preliminary experimentation is to observe a subject at some subjectively chosen dose. If there is no response, a second subject is observed at a higher dose; if the first subject responds, a second subject is observed at a lower dose. The preliminary experimentation proceeds in this sequential manner until the location is approximately determined. Perhaps only three or four total subjects are involved.

It should be noted that there are statistical estimation techniques based on such sequential design strategies (Anbar, 1977; Wu, 1985; Bruce, 1987). These sequential designs are of use when the allotted time for toxic response is brief and when experimental conditions can be held constant over time.

Because the best design configuration is greatly dependent on the choice of optimality criterion, on the estimator, and on the practical constraints, an in-depth discussion is not possible here. Some sources are Brown (1966, 1970), Ramsey (1972), Finney (1978), Tsutakawa (1980), Kuo (1983), Kooijman (1983), and Müller and Schmitt (1990).

In general, the dose levels should be wide enough to enclose most of the dose-response curve ($P(x_1)$ near 0 and $P(x_k)$ near 1) and centered as much as possible around the true location. In general, it is better to use a few subjects at many doses rather than many subjects at a few doses, other things being equal.

Engeman *et al.* (1986), performed a simulation study of small sample bioassays. They conclude that an absolutely minimum design involves at least 20 subjects observed at dose levels between the 5th and 95th percentiles of the true tolerance distribution. For smaller

samples, they state, "the resulting estimates lack credibility and should be considered at most a dose ranging study."

E. Confidence Interval for the LD50 When There Is Partial Mortality at Fewer Than Two Dose Levels

Not infrequently, an experimenter will choose such a wide dose mesh that there is partial mortality at none or only one of the dose levels. Confidence intervals associated with the nonparametric/robust procedures are probably not reliable for such coarse data. The standard maximum likelihood fit of a logistic or normal model will also fail to provide a confidence interval. Williams (1986) describes a likelihood ratio interval appropriate for the case where one can assume a parametric model, such as the logistic or normal.

References

- APHA (American Public Health Association), American Water Works Association, and Water Pollution Control Federation. (1981). *Standard Methods for the Examination of Water and Wastewater*. 15th ed. American Public Health Association, Washington, D.C.
- Altshuler, B. (1981). Modeling of dose-response relationships. *Environmental Health Perspectives* 42, 23-27.
- Anbar, D. (1977). The application of stochastic approximation methods to the bioassay problem. *Journal of Statistical Planning & Inference*. 1, 191-206.
- Aranda-Ordaz, F.J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* 68, 357-363.
- Azzalini, A., Bowman, A.W., and Hardle, W. (1989) On the use of nonparametric regression for model checking. *Biometrika* 76, 1-11.
- Baker, R.J. and Nelder, J.A. (1978). *The GLIM System, Release 3*. Numerical Algorithms Group, Oxford.
- Bennett, B.M. (1952). Estimation of the LD50 by moving averages. *Journal of Hygiene* 50, 157-164.
- Berkson, J. (1957). Tables for use in estimating the normal distribution function by normit analysis. *Biometrika* 44, 411-435.
- Bickel, P.J. and Lehmann, E.L. (1975). Descriptive statistics for nonparametric models: II. Location. *Annals of Statistics* 3, 1045-1069.
- Brown, B.W. (1961). Some properties of the Spearman estimator in bioassay. *Biometrika* 48, 293-302.
- Brown, B.W. (1966). Planning a quantal assay of potency. *Biometrics* 22, 322-329.
- Brown, B.W. (1970). Quantal-response assays. In: *Statistics in Endocrinology*. (J. W. McArthur and T. Colton, eds.) The MIT Press, Cambridge.

- Brown, C.C. (1982). On a goodness of fit test for the logistic model based on score statistics. *Communications in Statistics-Theory and Methods* 11, 1087-1105.
- Bruce, R.D. (1987). A confirmatory study of the up-and-down method for acute oral toxicity testing. *Fundamental and Applied Toxicology* 8, 97-100.
- Brunk, H.D. (1981). "Estimation of stimulus-response curves by Bayesian least squares" *Psychometrika* 46, 115-128.
- Chanter, D.O. (1975). Modifications of the angular transformation *Applied Statistics* 24, 354-359.
- Church, J.D. and Cobb, E.B. (1973). On the equivalence of Spearman-Kärber and maximum likelihood estimates of the mean. *Journal of the American Statistical Association* 68, 201-2.
- Cobb, E.B. and Church, J.D. (1983). Small-sample quantal response methods for estimating the location parameter for a location-scale family of dose-response curves. *Journal of the American Statistical Association* 78, 99-107.
- Copas, J.B. (1983). Plotting p against x . *Applied Statistics* 32, 25-31.
- Copenhaver, T.W. and Mielke, P.W. (1977). Quantal analysis: A quantal assay refinement. *Biometrics* 33, 175-186.
- Cornfield, J., Carlborg, F.W. and Van Ryzin, J. (1978). Setting tolerances on the basis of mathematical treatment of dose-response data extrapolated to low doses. In: *Proceedings of the First International Congress on Toxicology* (G.L. Plaa and W.A.M. Duncan, eds.). Academic Press, New York. pp. 143-164.
- Cramer, E.M. (1964). Some comparisons of methods of fitting the dosage response curve for small samples. *Journal of the American Statistical Association* 59, 779-793.
- Dalbey, W. and Lock, S. (1982). Inhalation toxicology of diesel fuel obscurant aerosol in Sprague-Dawley rats. ORNL/TM-8867, Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
- Disch, D. (1981). Bayesian nonparametric inference for effective doses in a quantal-response experiment. *Biometrics* 37, 713-722.
- Dixon, W.J., chief editor (1981). *BMDP Statistical Software—1981*. University of California Press, Berkeley.
- Doull, J., Klaassen, C.D. and Amdur, M.O. (1980). *Casarett and Doull's Toxicology: The Basic Science of Poisons*. 2nd ed. Macmillan Publ. Co., New York.
- Dykstra, R.L. (1981). An isotonic regression algorithm. *Journal of Statistical Planning and Inference* 5, 355-363.
- Engeman, R.M., Otis, D.L. and Dusenberry, W.E. (1986). Small sample comparison of Thompson's estimator to some common bioassay estimators. *Journal of Statistical Computation and Simulation* 25, 237-250.
- EPA: U.S. Environmental Protection Agency (1986). *Spearman/LC50 model: Trimmed Spearman-Kärber estimation of the LC50 values*. Version 1.0, Center for Water Quality Modelling, Environmental Research Laboratory, College Station Road, Athens, GA 30613.
- Finney, D.J. (1971). *Probit Analysis*, 3rd ed. Cambridge University Press, London.
- Finney, D.J. (1978). *Statistical Method in Biological Assay*. 3rd ed. Macmillan Publ. Co., New York.
- Fowlkes, E.B. (1987) Some diagnostics for binary logistic regression via smoothing. *Biometrika* 74, 503-515.

- Goldstein, A., Aronow, L. and Kalman, S.M. (1974). *Principles of Drug Action: The Basis of Pharmacology*. 2nd ed. John Wiley and Sons, New York.
- Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response models. *Biometrika* 69, 309-314.
- Hamilton, M.A. (1977). Estimating the logistic curve from grouped data. *Journal of Statistical Computation and Simulation* 5, 279-301.
- Hamilton, M.A. (1979). Robust estimates of the ED50. *Journal of the American Statistical Association* 74, 344-354.
- Hamilton, M.A. (1980). Inference about the ED50 using the Trimmed Spearman-Kärber procedure—A Monte Carlo study. *Communications in Statistics—Simulation and Computation* B9, 235-254.
- Hamilton, M.A., Russo, R.C. and Thurston, R.V. (1977). Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. *Environmental Science and Technology* 11, 714-719. Correction: 12, 417.
- Harris, E.K. (1959). Confidence limits for the LD50 using the moving average-angle method. *Biometrics* 15, 424-432.
- Hasselblad, V., Stead, A.G. and Creason, J.P. (1980). Multiple probit analysis with a nonzero background. *Biometrics* 36, 659-663.
- Hewlett, P.S. and Plackett, R.L. (1979). *The Interpretation of Quantal Responses in Biology*. Edward Arnold (Publishers) Limited, London.
- Hoel, D.G. (1980). Incorporation of background in dose-response models. *Federation Proceedings* 39, 73-75.
- Huber, P.J. (1981). *Robust Statistics*. John Wiley and Sons, New York.
- Hubert, John J. (1984). *Bioassay*. Kendall/Hunt Publishing Company, Toronto.
- Hunter, W.G. and Crowley, J.J. (1979). Hazardous substances, the environment and public health: A statistical overview. *Environmental Health Perspectives* 32, 241-254.
- Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation. *Annals of the Institute of Statistical Mathematics* 35, sect. B, 115-137.
- Jacquez, J.A. (1972). *Compartmental Analysis in Biology and Medicine*. Elsevier, New York.
- James, B.R. and James, K.L. (1983). On the influence curve for quantal bioassay. *Journal of Statistical Planning and Inference* 8, 331-345.
- James, B.R., James, K.L. and Westenberger, H. (1984). An efficient R-estimator for the ED50. *Journal of the American Statistical Association* 79, 164-173.
- Jennings, D.E. (1986). Outliers and residual distributions in logistic regression. *Journal of the American Statistical Association* 81, 987-990.
- Kappenman, R.F. 1987. Nonparametric estimation of dose-response curves with application to ED50 estimation *Journal of Statistical Computation and Simulation* 28, 1-13.
- Kooijman, S.A.L.M. (1983). Statistical aspects of the determination of mortality rates in bioassays. *Water Research* 17, 749-759.
- Kuo, L. (1983). Bayesian bioassay design. *Annals of Statistics* 11, 886-895.
- Lemeshow, S. and Hosmer, D.W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology* 115, 92-106.

- Litchfield, J.T. and Wilcoxon, F. (1949). A simplified method of evaluating dose-effect experiments. *Journal of Pharmacology and Experimental Therapeutics* 96, 99-113.
- Mancini, J.L. (1983). A method for calculating effects, on aquatic organisms, of time varying concentrations. *Water Research* 17, 1355-1362.
- Matsumura, F. (1985). *Toxicology of Insecticides*. Plenum Press, New York.
- McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*. Chapman and Hall, London.
- Miller, R.G. (1973). Nonparametric estimators of the mean tolerance in bioassay. *Biometrika* 60, 535-542.
- Miller, R.G. and Halpern, J.W. (1980). Robust estimators for quantal bioassay. *Biometrika* 67, 103-110.
- Morton, R. (1981). Generalized Spearman estimators of relative dose. *Biometrics* 37, 223-233.
- Müller, H. and Schmitt, T. (1988) Kernel and probit estimates in quantal bioassay. *Journal of the American Statistician Association* 83, 750-759.
- Müller, H. and Schmitt, T. (1990) Choice of number of doses for maximum likelihood estimation of the ED50 for quantal dose-response data. *Biometrics* 46, 117-129.
- Murray, G.D. (1983). Nonconvergence of the minimax order algorithm. *Biometrika* 70, 490-491.
- National Research Council (1984). *Toxicity Testing: Strategies to Determine Needs and Priorities*. National Academy Press, Washington, D.C.
- Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics* 38, 485-498.
- Prentice, R.L. (1976). A generalization of the probit and logit methods for dose-response curves. *Biometrics* 32, 761-768.
- Racine, A., Grieve, A.P. and Fluhler, H. (1986). Bayesian methods in practice: Experiences in the pharmaceutical industry. *Applied Statistics* 35, 93-150.
- Ramsey, F.L. (1972). A Bayesian approach to bioassay. *Biometrics* 28, 841-858.
- Sanathanan, L., Gade, E. and Shipkowitz, N. (1987). Trimmed logit method for estimating the ED50 in quantal bioassay. *Biometrics* 43, 825-832.
- SAS Institute Inc. (1982). *SAS User's Guide: Statistics—1982 Ed.* SAS Institute, Inc., Cary, N.C.
- Schmoyer, R.L. (1984). Sigmoidally constrained maximum likelihood estimation in quantal bioassay. *Journal of the American Statistical Association* 79, 448-453.
- 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.
- Sprague, J.B. (1969). Measurement of pollutant toxicity to fish: I. Bioassay methods for acute toxicity. *Water Research* 3, 793-821.
- Staniswalis, J.G. and Cooper, V. (1988) Kernel estimates of dose response. *Biometrics* 44, 1103-1119.
- Stephan, C.E. (1977). Methods for calculating an LC50. In: *Aquatic Toxicology and Hazard Evaluation* (ASTM STP 634). (F.L. Mayer and J.L. Hamelink, eds.). American Society for Testing and Materials, Philadelphia.
- Stewart, L. (1979). Multiparameter Univariate Bayesian Analysis. *Journal of the American Statistical Association* 74, 684-693.

- Tsutakawa, R.K. (1980). Selection of dose levels for estimating a percentage point of a logistic quantal response curve. *Applied Statistics* 29, 25-33.
- Tsutakawa, R.K. (1982). Bioassay, Statistical Methods. In: *Encyclopedia of Statistical Sciences*, v.1. (S. Kotz and N.L. Johnson, eds.). John Wiley and Sons, New York.
- Warren, C.E. (1971). *Biology and Water Pollution Control*. W.B. Saunders Co., Philadelphia.
- Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* 31, 144-148.
- Williams, D.A. (1986). Interval estimation of the median lethal dose. *Biometrics* 42, 641-645.
- Wolfe, R.A., Roi, L.D. and Margosches, E.H. (1981). Monotonic dichotomous regression estimates: A burn care example. *Biometrics* 37, 157-167.
- Wright, F.T. (1978). Estimating strictly increasing regression functions. *Journal of the American Statistical Association* 73, 636-639.
- Wu, C.F.J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association* 80, 974-984.