

# Small Sample Comparison of Thompson's Estimator to Some Common Bioassay Estimators

RICHARD M. ENGEMAN, DAVID L. OTIS and WILLIAM E. DUSENBERRY

*Denver Wildlife Research Center, U.S. Fish and Wildlife Service,  
Building 16, Denver Federal Center, Denver, Colorado 80225, U.S.A.*

*(Received March 1, 1985; in final form February 12, 1986)*

For a variety of reasons, bioassay experiments are frequently conducted with small sample sizes. However, the statistical literature offers little comparative information for small sample sizes among the analytical procedures most typically applied in vertebrate studies for chemical registration. A simulation study was performed to compare several traditional analytical procedures for estimating the median lethal dose ( $LD_{50}$ ). These estimation procedures include probit analysis with maximum likelihood estimation, logit analysis with maximum likelihood and minimum chi-square estimation methods, and analysis by a nonparametric moving average procedure. The simulation results indicate the conditions under which each analytical method appears most useful. Recommendations are tentatively made for minimally adequate designs for bioassay studies.

**KEY WORDS:** Bioassay, small sample, probit, logit, moving average,  $LD_{50}$ .

## 1. INTRODUCTION

Little information is available on the small sample properties of the procedures for analyzing the data from bioassay experiments. Such bioassay studies with small sample sizes are particularly common in

wildlife biology and related fields, for a variety of reasons. The subject animals may be rare, difficult to capture, or expensive to buy. Governmental regulations such as quarantine procedures or permit limits may also restrict sample sizes. Unfortunately, some investigators use small samples simply because an analytical technique may require only a small amount of data to yield an estimate, even though the small sample properties of that technique are unknown. For situations when only a small number of subjects are available, information about the properties of the various analytical methods would be useful for developing and justifying a study protocol.

The few recommendations that are available on minimum sample sizes and analytical techniques are inconsistent. In the avian and mammalian testing section of EPA's Proposed Guidelines (1978) for the Registration of Pesticides in the United States, the recommended study design calls for at least 5 dose levels, plus a control, and at least 10 animals per level to estimate an avian median lethal (single) dose  $LD_{50}$ . The suggested estimation method is described as "any acceptable method". Only two estimation methods are specifically mentioned as acceptable: those due to Thompson (1947) and to Litchfield and Wilcoxon (1949). However, the suggested study design for estimating a dietary median lethal concentration  $LC_{50}$ , one page later in the same document, calls for 6 concentration levels with at least 6 animals per level. The recommended estimation method is "probit analysis as described by Finney (1971)". The American Society for Testing and Materials (ASTM) (1976) suggests that 12 animals, 6 male and 6 female, should be used. No minimum number of doses nor minimum number of animals per dose is indicated. Again, the two estimation procedures specifically mentioned as acceptable are those due to Litchfield and Wilcoxon (1949) and to Thompson (1947), although publications by Finney (1971) and Bliss (1938) are mentioned as containing "other equally reliable methods". The motivations behind the above recommendations for design and estimation are not given. In contrast to these recommendations on a minimal experimental design, Finney (1971) expressed mild reservations concerning the possible bias in parametric estimates from experiments as large as 100 animals in each of 8 dose levels.

Similarly, the statistical literature contains little information comparing the small sample properties of the various analytical techniques for bioassay experiments when small sample numbers are

unavoidable. Most analytical procedures for estimating quantal responses and associated confidence intervals rely on asymptotic theoretical results or approximations to establish their properties. The problem of determining finite sample properties, particularly for small samples, has long been recognized (e.g., Berkson, 1955a). Sowden (1971) demonstrated that, for quantal response models, estimates can be quite biased and their asymptotic variances in substantial error. Cramer (1964) compared three estimation methods for fitting the normal integrated response curve, but did not consider comparing other analytical procedures such as the use of logits or moving averages. Hamilton (1979) compared 10 nonparametric and logit estimators in moderately-sized experiments with 10 dose levels and 5 to 20 individuals per level. Cobb and Church (1983) studied the small sample properties of several estimation methods for a general family of dose-response curves, but did not compare some of the traditional analytical methods in use such as probit or nonparametric methods. Smith *et al.* (1984) compared maximum likelihood and minimum chi-square techniques for logit analyses with moderate sample sizes. Oglesby and Bundrick (1981) compared several analytical approaches for estimating an  $LD_{50}$ , including probit analysis and Thompson's (1947) moving average method, but they were investigating primarily moderate sample sizes. They suggested that when distributional assumptions are not satisfied, the Thompson method should be given consideration.

In light of the irresolution among the recommendations and the lack of comparative information among analytical procedures, we studied the effects of small sample sizes and small numbers of dose levels on the estimates arrived at through the analytical procedures most commonly applied and recommended in the literature pertaining to vertebrate studies for registration of chemicals. In particular, we were interested in how well the often used and recommended procedure of Thompson (1947) compared with the most commonly used parametric methods.

## 2. ANALYTICAL PROCEDURES

In addition to Thompson's (1947) nonparametric method, the probit and logit methods were the parametric analyses considered. Only

maximum likelihood estimation was considered for the probit method; however, for the logit analyses we considered a minimum chi-square estimation technique in addition to that of maximum likelihood. Since these methods are well documented in the literature, only a brief description and references for each are given.

The probit method of bioassay analysis was initiated by Bliss (1935a, 1935b, 1938) and further developed by Finney (1971). The theory underlying probit analysis is that the hypothetical frequency distribution of susceptibility as measured by individual lethal doses is normally distributed. In other words, the dose-response curve follows the cumulative normal frequency distribution. Since this curve is difficult to work with, Bliss (1935a) developed the probit transformation. This transformation effectively straightens the sigmoid dose response curve and allows the  $LD_{50}$  and, in fact, the complete range of LD values to be estimated based on a weighted linear regression model.

The logit method of bioassay analysis (Berkson, 1944) is similar to the probit method except that the underlying tolerance distribution is assumed to be logistic rather than normal (see, for example, Berkson, 1944, 1951; Finney, 1971). The logistic distribution is a bell-shaped curve similar to the normal curve, but it has heavier tails, i.e., the probabilities of extreme values are larger. However, it is well known (e.g., Armitage and Allen, 1950; Berkson, 1950; Miller, 1950; Finney, 1971) that probit and logit methods often result in estimates of LD values that are very close and often indistinguishable in moderate to large sample sizes. Instead of calculating probits for the dose-effect regression line, logit values are produced by the logit transformation (e.g., Finney, 1971) and the linear regression model is solved by either maximum likelihood or minimum chi-square techniques (e.g., Cox, 1970).

Many early biologists had a preference for a parametric analysis because they wanted information on the whole range of lethal dose (LD) values (e.g., from  $LD_{10}$  to  $LD_{90}$ ). Many also believed that the assumed underlying response distribution for their analysis was appropriate. However, the tedious mathematical calculations of the parametric methods and the lack of efficient computing often precluded their use. Also, some were aware that parametric methods for estimating LD values may not be robust to the failure of the assumption concerning the form of the tolerance distribution for the population of animals being considered.

Potential bias caused by lack of fit of the data to the underlying distribution and the computational difficulties for the parametric estimators led Thompson (1947) to develop a distribution-free, moving average method for calculating the median lethal dose ( $LD_{50}$ ). Although there are a number of nonparametric estimators available in the literature (e.g., Finney, 1971), we study the Thompson estimator because it is the standard procedure in the environmental protection literature. The Thompson method uses moving averages on the proportion of animals responding at each dose level followed by interpolations to arrive at an estimate of the  $LD_{50}$ .

The Thompson method has certain advantages over the probit and logit methods. It is nonparametric, relatively simple to calculate, and it can produce an estimate when there are as few as two dose levels of two animals each. Disadvantages to the Thompson method include the fact that only the  $LD_{50}$  and its approximate standard error can be computed, and that all dose levels must be equally spaced on a geometric scale.

### 3. SIMULATION SETUP

Evaluation of the relative performance of the estimation methods was done through a Monte Carlo simulation study. Five bioassay estimators were considered for the simulation: probit maximum likelihood (ML), logit ML, logit minimum chi-square (MCS), and the Thompson moving average method.

The first of two FORTRAN programs used in the study generated the data and performed the computations necessary for the estimation procedures. Most of this program was developed by the U.S. Forest Service, Pacific Southwest Forest and Range Experiment Station. The probit and logit procedures from their published program, POLO (Russell *et al.*, 1977), were incorporated into the simulation program. These researchers performed their own simulation study on bioassay methods, but were concerned with response of insects to insecticides and, therefore, were working with sample sizes much greater than those usually available in vertebrate studies (Smith *et al.*, 1984). The minimum chi-square method that we used was the two-stage procedure of Cox (1970), although Smith *et al.* (1984) subsequently used the MCS method of Berkson (1955b) in

their study. Thompson's moving average method was the only major addition to their program.

Because the Thompson method can be used only with geometrically spaced dose levels, the simulation study was restricted to experiments of this type. The simulated data were generated from a theoretical population following a logistic distribution with location parameter equal to five and scale parameter equal to one. Thus, for our simulations, the true  $LD_{50}$  was equal to 5.

The basic configuration for the dose levels consisted of equally spaced doses on a geometric scale with the theoretical levels of response for the first and last doses set at 0.05 and 0.95, respectively. This pattern was applied to experiments with 3, 4, 5 and 6 dose levels. We simulated these experiments using 2, 4, 6, 10 and 20 subjects per dose level. For each experimental combination (number of doses by number of subjects per dose), 5000 replications were simulated.

We developed a second FORTRAN program, to summarize the results of the simulations into statistics for comparing and evaluating the estimation methods. For each replication of each experiment, the essential results generated for each estimation procedure were the estimated  $LD_{50}$  and its associated 95 percent confidence intervals. The statistics calculated in the summary program included estimated mean squared error, bias, and variance for each estimator of the  $LD_{50}$ . Also calculated was confidence interval coverage, that is, the percentage of replications in which the calculated 95 percent confidence interval included the true LD value. The average width of the 95 percent interval was also calculated. For our own information, these estimates and statistics were also calculated for estimating the  $LD_{90}$  using the parametric estimators. If any estimator failed to calculate an estimate in a given replication of the simulation, then none of the results for any of the estimators were allowed in the calculation of the summary statistics (failure rates for most simulation setups were less than 25%). Thus, each replication that is included in the summarization can be considered a member of an ideal (conditional) set of responses where all estimators are able to produce finite point and variance estimates. We do not consider in this study the much more complex situation for investigating the properties of each of the estimators when one or more of the others fail.

#### 4. SIMULATION RESULTS

The mean squared error (MSE) results from simulations are given in Table I. As would be expected, the MSE for each estimator almost always decreased as the number of subjects per dose level increased and/or as the number of dose levels increased. For estimating the  $LD_{50}$ , the probit ML and the two logit estimators, in general, had similar MSE's. For the cases with 6 or fewer subjects per dose level, the logit MCS estimator tended to have a slightly smaller MSE than the probit or logit ML (with the logit MCS having the smallest MSE in each case). In every dose level by sample size combination, the Thompson moving average method had the largest MSE.

TABLE I  
Mean squared error for estimators of  $LD_{50}$

No. of dose levels	No. of animals/level	Probit ML	Logit ML	Logit MCS	Thompson
3	4	0.665	0.620	0.447	0.744
	6	0.429	0.410	0.314	0.478
	10	0.282	0.279	0.233	0.312
	20	0.160	0.166	0.147	0.180
4	2	0.445	0.434	0.234	0.791
	4	0.452	0.443	0.358	0.560
	6	0.344	0.341	0.301	0.438
	10	0.219	0.217	0.208	0.282
	20	0.117	0.116	0.116	0.167
5	2	0.459	0.456	0.380	0.683
	4	0.359	0.361	0.319	0.446
	6	0.276	0.278	0.263	0.342
	10	0.169	0.173	0.169	0.230
6	2	0.088	0.090	0.090	0.135
	4	0.512	0.513	0.423	0.671
	6	0.302	0.335	0.330	0.409
	10	0.226	0.229	0.216	0.282
	20	0.143	0.146	0.144	0.191
	20	0.071	0.072	0.072	0.112

The results for the magnitude of the biases of the  $LD_{50}$  estimates (Table II) parallel those for the MSE's. The Thompson method had the largest absolute bias in every case whereas biases of the

TABLE II  
Bias for estimators of LD<sub>50</sub>

No. of dose levels	No. of animals/level	Probit ML	Logit ML	Logit MCS	Thompson
3	4	-0.505	-0.501	-0.356	-0.763
	6	-0.247	-0.253	-0.124	-0.504
	10	-0.075	-0.094	+0.014	-0.313
	20	+0.028	-0.014	+0.067	-0.216
4	2	-0.648	-0.638	-0.434	-0.754
	4	-0.210	-0.203	-0.092	-0.397
	6	-0.109	-0.105	-0.012	-0.278
	10	-0.023	-0.027	+0.045	-0.197
5	2	-0.335	-0.323	-0.197	-0.477
	4	-0.055	-0.050	+0.027	-0.206
	6	-0.005	-0.005	+0.057	-0.164
	10	-0.002	-0.004	+0.047	-0.159
6	2	+0.013	+0.006	+0.038	-0.165
	4	-0.237	-0.221	-0.122	-0.341
	6	-0.012	-0.007	+0.055	-0.140
	10	-0.003	-0.002	+0.049	-0.156
	20	+0.010	+0.008	+0.046	-0.148
	20	+0.004	-0.003	+0.022	-0.170

parametric estimators were similar, except at the smallest sample sizes, where the logit MCS had smallest absolute bias. For every dose level by sample size combination considered, the Thompson estimator had a negative bias (underestimated the true LD<sub>50</sub>).

Table III contains the results for the percent of replications in which the 95% confidence interval actually contained the true value being estimated. Coverage achieved or exceeded the stated (95%) level for each of the two logit methods and the probit method except for the cases of 2 animals per dose (regardless of the number of dose levels) and 4 animals per dose when there were 3 dose levels. In these instances, the Thompson method exhibited higher coverage of the true LD<sub>50</sub> although it did not achieve the stated 95% level. Generally, the Thompson method did not compare favorably to the other methods. Because the Thompson method usually achieved only 84% to 93% coverage, its variance approximation may be negatively

TABLE III  
 Percentage coverage of actual LD<sub>50</sub> by 95% confidence intervals  
 (conditional set of replications where no estimator failed)

No. of dose levels	No. of animals/level	Probit ML	Logit ML	Logit MCS	Thompson
3	4	93.8	62.5	85.0	100.0
	6	99.5	95.2	99.1	96.7
	10	99.9	99.9	99.9	91.9
	20	98.3	99.4	99.8	91.9
4	2	43.4	43.4	43.4	93.2
	4	98.5	97.4	97.9	93.2
	6	99.8	99.9	99.7	90.1
	10	96.8	96.8	98.5	90.6
5	2	59.4	59.4	59.4	84.2
	4	99.6	99.3	99.3	89.4
	6	97.9	98.1	98.8	90.3
	10	96.2	96.8	97.4	89.1
6	2	78.5	51.1	68.9	71.0
	4	98.6	99.2	99.2	86.5
	6	96.6	97.4	98.6	89.4
	10	95.3	96.0	96.7	89.2
	20	94.9	95.2	95.9	86.7

biased (underestimation of variance) or it requires a much larger sample size before it becomes accurate. The consistent negative bias in the LD<sub>50</sub> estimate itself could also decrease coverage.

The results on confidence interval width (Table IV) were necessarily associated with the results on confidence interval coverage. In most designs for estimating the LD<sub>50</sub>, the Thompson method resulted in the narrowest 95% confidence intervals; however, few of these designs also produced 95% coverage of the true LD<sub>50</sub> using the Thompson method. As indicated above, this was probably due to a variance approximation that is too small. In most simulations, the probit ML method confidence intervals were next narrowest, followed by logit ML and then the logit MCS estimator.

Although the Thompson method did not perform as well as the

TABLE IV  
Mean width of 95% confidence intervals on LD<sub>50</sub>

No. of dose levels	No. of animals/level	Probit ML	Logit ML	Logit MCS	Thompson
3	4	127.400	2.090	58.140	4.285
	6	5.955	7.465	10.730	3.205
	10	3.155	3.586	4.117	2.295
	20	1.813	2.025	2.169	1.559
4	2	1.618	1.618	1.618	3.476
	4	6.990	11.310	15.090	2.944
	6	4.025	5.162	6.995	2.403
	10	2.219	2.382	2.778	1.858
5	20	1.408	1.457	1.601	1.326
	2	1.663	1.663	1.663	2.587
	4	5.497	7.570	13.050	2.379
	6	2.862	3.078	4.225	2.042
6	10	1.830	1.932	2.188	1.595
	20	1.208	1.248	1.339	1.165
	2	20.610	1.534	12.540	2.071
	4	3.848	4.269	14.260	2.053
	6	2.428	2.665	3.321	1.815
	10	1.614	1.690	1.877	1.449
	20	1.073	1.106	1.171	1.049

others judging by most of the above criteria, it was clearly the best with respect to failure rate since it consistently was able to produce an LD<sub>50</sub> estimate more often than the other procedures. This difference was particularly pronounced when the total number of subjects was less than 20. For example, the failure rate for the parametric methods was as high as 70% for the case of 3 dose levels with 4 animals per level. In these cases, the Thompson method usually had a failure rate one-third to one-fifth that of the parametric methods. As the sample size and/or the number of dose levels increased, the failure rate for each method declined. For the larger sample sizes (e.g., more than 3 dose levels and at least 30 subjects overall), the failure rates for calculating an LD<sub>50</sub> were negligible (0-7%) for each method.

## 5. DISCUSSION

Although there is a considerable body of literature to ponder when choosing an analytical procedure to use for a given bioassay, most of this information relates to moderate or large sample size situations. Little information is available to help decide what procedure to use for the small sample size situation. If a study is conducted with small sample sizes, a goodness-of-fit test would be of little use for selecting among similar parametric tolerance distributions such as probit or logit analyses. Also, the particular computational method must be selected, e.g., maximum likelihood versus minimum chi-square. A nonparametric procedure may be suggested to avoid distributional concerns and for ease of calculation, but at the cost of less information produced and more prior constraints on the experimental design. In this vein, Finney (1971, p. 40) maintained that satisfactory nonparametric methods had not been reported.

If the study is designed with the Thompson method in mind, one must consider its performance relative to that of competing estimators. In general, the Thompson method did not perform as well as the other four methods studied for estimating the  $LD_{50}$ . In particular, if underestimation of the  $LD_{50}$  (i.e., overestimation of toxicity) can have seriously adverse consequences, then the Thompson method should be given less consideration. However, the failure rate of the Thompson method was substantially less for the simulation involving smaller sample sizes. This property could represent a real advantage if the Thompson method was able to produce reasonable estimates when other methods fail. For this reason, we ran further simulations on only the Thompson method for two designs in which the failure rates for the parametric methods were about three times greater than that of the Thompson method (38 and 33% versus 11 and 12%). Results for the design having 4 dose levels and 4 subjects per level were acceptable since they were very similar to those reported for the Thompson method in Tables I-IV. However, for the design having 6 dose levels and 2 subjects per level, performance of the Thompson method was poor: e.g., the confidence interval coverage was only 58 percent. Finney (1951) indicates that the Thompson method's greatest advantage is its admissibility of a limited dose range (the range needs only to be wide enough to include the  $LD_{50}$ ), but if the experimenter has enough previous

knowledge, then a more efficient experiment can be designed with other analytical techniques in mind. More study is needed to further define when the Thompson method can produce useful estimates if the data are insufficient for the more rigorous parametric techniques. It should also be remembered that we only considered a symmetric underlying distribution (logistic), and we would, therefore, expect the logit and probit estimators to perform well. Perhaps with a more unusual (e.g., asymmetric) underlying distribution, the nonparametric Thompson method would perform better relative to the other four methods (e.g., Oglesby and Bundrick, 1981).

For other than the very small sample sizes, the parametric methods outperform the Thompson method for estimating the  $LD_{50}$  and their failure rates are not much greater than that for the Thompson method. With the current availability and technology of computers and calculators, computational ease is no longer a strong reason for considering the Thompson method over these parametric methods.

Based on our study, we suggest the following as an absolutely minimum design for a bioassay study: at least four dose levels and *more* than four subjects per dose level; otherwise the resulting estimates would lack credibility and should be considered to be at most a dose ranging study. This minimal design should be considered only if the doses can be at least approximately balanced similar to the simulation setup described in Section 3. For dose ranging studies, it appears that the Thompson estimation method would be of most use (see also Finney, 1951). However, for studies satisfying at least the above minimal recommendations, we suggest the use of the probit or logit ML procedures.

Before more general recommendations can be set forth on small sample design and estimation methods, more extensive study on the effects of the underlying tolerance distribution must be done. Additionally, studies on the effects of where the doses lie in the tolerance distribution should be done (Hertzberg, 1975 and Smith *et al.*, 1984, indicated that dose allocation can significantly affect the performance of estimators). We ran some additional simulations in which the theoretical responses for the doses were skewed, rather than the balanced pattern discussed here. In these few probes, the relative results for the estimators remained the same as described in Section 4. Also, further investigation of the properties of the other

estimators when one estimator fails would allow stronger recommendations on study design and analysis.

This study was restricted to the bioassay estimation techniques most widely recommended and applied in the literature pertaining to vertebrate studies for the registration of chemicals. The statistical literature contains variations on these techniques as well as other techniques which may offer improvements. Some of these methods may provide estimates for even smaller sample sizes (e.g., a binomial test, Siegel, 1956) or they may be more efficient. However, to maintain the study at a tractable size and within computing resources, we chose to limit the study only to these most widely practised methods.

Bioassay studies with small sample sizes will become more prevalent as more studies are done on rare species, as research dollars become scarcer, and as the animal welfare movement becomes more effective at inhibiting the use of animals in research (e.g., Holden, 1982). The present study provides the researcher with useful information on design and data analysis considerations for bioassay studies where small sample sizes are unavoidable.

### Acknowledgements

We are grateful to Thomas Keefe and Edward W. Schafer, Jr. for their helpful reviews of the paper. Valuable programming and computing assistance were provided by Steve Hotovy, Gary Hahnenkamp and Gary Walkovitz of Boeing Computer Services Company. Significant initial discussions and software were provided by Kim Smith and Jackie Robertson of the U.S. Forest Service, Berkeley, California. This project was supported, in part, by Interagency Agreement No. 14-16-0009-81-957 funded by the U.S. Environmental Protection Agency.

### References

- Armitage, P. and Allen, I. (1950). Methods of estimating the  $LD_{50}$  in quantal response data. *J. of Hygiene* **48**, 298-322.
- ASTM Standards on Vertebrate Control Agents. (1976). American Society for Testing and Materials, Philadelphia, PA.
- Berkson, J. (1944). Application of the logistic function to bio-assay. *J. Amer. Stat. Assoc.* **39**, 357-365.
- Berkson, J. (1950). Some observations with respect to the error of bio-assay. *Biometrics* **6**, 432-434.
- Berkson, J. (1951). Why I prefer logits to probits. *Biometrics* **7**, 327-339.
- Berkson, J. (1955a). Estimate of the integrated normal curve by minimum normit and chi-square with particular reference to bio-assay. *J. Amer. Stat. Assoc.* **50**, 529-549.

- Berkson, J. (1955b). Maximum likelihood and minimum  $X^2$  estimates of the logistic function. *J. Amer. Stat. Assoc.* **50**, 130–162.
- Bliss, C. L. (1935a). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134–167.
- Bliss, C. L. (1935b). The comparison of dosage-mortality data. *Ann. Appl. Biol.* **22**, 307–333.
- Bliss, C. L. (1938). The determination of the dosage-mortality curve from small numbers. *Quart. J. Pharm. Pharmacol.* **11**, 192–216.
- Cobb, E. B. and J. D. Church. (1983). Small-sample quantal response methods for estimating the location parameter for a location-scale family of dose-response curves. *J. Amer. Stat. Assoc.* **78**, 99–107.
- Cox, D. R. (1970). *The Analysis of Binary Data*. Chapman and Hall, London.
- Cramer, E. M. (1964). Some comparisons of methods of fitting the dosage response curve for small samples. *J. Amer. Stat. Assoc.* **59**, 779–793.
- Environmental Protection Agency. (1978). Avian and mammalian testing. Registration of pesticides in the United States proposed guidelines. *Federal Register* **43**, 29726–29734.
- Finney, D. J. (1951). The estimation of the mean of a normal tolerance distribution. *Sankhya* **10**, 341–360.
- Finney, D. J. (1971). *Statistical Method in Biological Assay*. Griffin–Cambridge Univ. Press, London.
- Hamilton, M. A. (1979). Robust estimates of the ED50. *J. Amer. Stat. Assoc.* **74**, 344–354.
- Hertzberg, J. O. (1975). On small sample properties of probit analysis. In: *Proceedings of the 8th International Biometric Conference*, L. C. A. Corsten and T. Postelnicu (eds.) **8**, 153–162.
- Holden, C. (1982). New focus on replacing animals in the lab. *Science* **215**, 35–38.
- Litchfield, J. T. and Wilcoxon, F. (1949). A simplified method of evaluating dose-effect experiments. *J. Pharmacology and Experimental Therapeutics* **96**, 99–113.
- Miller, L. C. (1950). Biological assays involving quantal responses. *Annals of the New York Academy of Sciences* **52**, 903–919.
- Oglesby, J. L. and Bundrick, C. M. (1981). A comparison of several approaches for calculating  $LC_{50}$ 's for binary responses obtained from toxicological experiments. *Proceedings of the Sixth Annual SAS User's Group International Conference*, 142–149.
- Russell, R. M., Robertson, J. L. and Savin, N. E. (1977). POLO: A new computer program for probit analysis. *Bull. Entomol. Soc. of Amer.* **23**, 209–213.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- 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.
- Sowden, R. R. (1971). Bias and accuracy of parameter estimates in a quantal response model. *Biometrika* **58**, 595–603.
- 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.