THE EFFECT OF RANDOM DOSAGES ON PROBIT ANALYSIS

by

BRUCE MORRILL

B. A., Kansas State University, 1977
M. A., Harding Graduate School, 1981

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1984

Approved by:

James J. Higgins

Major Professor

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	The Method of Probit Analysis	4
III.	The Design and Results of a Simulation	14
IV.	Conclusion	77
	References	80
	Acknowledgements	81

I. Introduction

A wide range of disciplines use dose-response statistical techniques in analyzing the relationship between levels of a stimulus and the resulting responses. Dose-response methodology has been primarily developed by researchers whose interest was to set the proper levels of pesticide required to kill a certain insect species. In biological research, the interest may be in determining the quantity of a drug needed to cure a disease, or the amount of a vaccine needed to prevent one. The dose may also represent time, as, for example, studies on the relation of age of subjects to certain biological functions or the amount of time exposed to a stimulus before a physical or chemical reaction occurs have used dose-response methods. Dose-response methods have been used with psychological experiments to investigate the relation of levels of measurable incentives to a decision making process. Engineering processes may be studied to determine how much disturbance a system or product can tolerate before it fails.

The application of dose-response methods presumes that the level of stimulus affects the nature of the response. If specified levels of the dose do not always result in exactly the same response, estimation of the relationship between dose and response by experimentation is necessary. In this case the varying responses to given dose levels are caused by variability in the test subjects. Moreover, additional variability in the estimated relationship will arise from factors which are present in a given

experiment and which are not controlled from experiment to experiment. Statistical methods are then necessary which both estimate the dose-response relationship given variability between subjects, and which allow for the application of experimental design techniques to reduce variability within and between experiments. The development of dose-response methodology has been directed to meeting these criteria.

Although many responses may be measured on a quantitative scale, many important questions concern a relationship where there are only two possible responses: either the subject responds or it does not. In the examples given above, either the disease is cured or it is not, or the disease is prevented or it is caught, or the insect is alive or it is as good as dead. These cases of 'either-or' or 'quantal' responses constitute a major portion of the dose-response methodology.

In a quantal response test, an experimental subject will be given a specified dose of the stimulus. Each subject is assumed to have a certain tolerance to the stimulus. Whether a subject responds or fails to respond will depend on whether the dosage given was above or below that subject's tolerance. The level of tolerance presumably varies from subject to subject, thus the tolerances of the population of subjects have a specific distribution. A dose-response model can be fit to this tolerance distribution and its parameters may be estimated using the experimental data.

A common quantal response analysis assumes that the underlying distribution of the tolerances is a normal. The analysis of this model is

called Probit Analysis. In Probit Analysis, the dose-response model is given by the following equation:

$$P(\text{response}\,|\,\text{dose}=x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(v-\mu)^2}{2\sigma^2}\right] dv$$

In this case, the variation in the responses of the subjects is binomial, depending upon whether the specified dosage given to a subject is above or below that subject's tolerance.

The interest of this paper is to determine the effects of random variation in the dosages on the results of Probit Analysis. The usual Probit Analysis assumes that the administered dosages are known and fixed. For example, if an experiment is conducted to determine the median effective dose of an insecticide, each of 5 groups of subjects may be exposed to one of 5 different dosage levels. It is assumed that each subject within a group is exposed to exactly the same amount of the insecticide. In many cases the quality of a specified concentration of insecticide and the care taken to ensure a uniform application of the insecticide to the subjects will be sufficient to justify this assumption. If, however, there is some doubt about the accuracy of an insecticide's concentration or about the method of application of that insecticide, there may also be subsequent doubts about the validity of the results of Probit Analysis. The intent of this paper is to give guidelines for the amount and nature of random dosage variation which can be tolerated before significant effects on the results of Probit Analysis are noted.

II. The Method of Probit Analysis

Dose-response techniques involve the application of a stimulus to a subject to achieve a desired response. The stimulus may be a pesticide, a drug, or time. The subject in these cases might be an insect, a person, or a process. The response, respectively, would be the death of the insect, the recovery of the patient, or the disruption of the process.

The dose-response techniques assume that there is a functional relationship between the level or dose of the stimulus and the nature or magnitude of the response. That is, when the dosage level is increased, the response rate is presumed to also increase (or decrease). Thus, if a drug is being tested in hopes of decreasing the blood pressure of a patient, it may be presumed that the percentage decrease in the blood pressure will depend on how much of the drug is administered.

The response is a measurement made on a specified characteristic of the subject. In many cases the response may be accurately measured on a quantitative scale, such as the percentage decrease in blood pressure. In other cases, a quantal or either-or response is involved. Quantal responses may be recorded either when quantitative measurements cannnot be accurately taken (e.g., the headache is either the same or better) or when the nature of the response is that it occurred or did not occur (the subject either lived or died).

The widespread discussion of dose-response techniques for both quantitative and quantal responses has been compiled in a book by Finney

(1978). The present study is limited to the quantal response problem, to which Finney (1971) had devoted an earlier book. The following discussion of the terminology and method of probit analysis is derived primarily from these two books.

When the response of a subject is an either-or or quantal response, it is presumed that the dosage level which elicits the desired response varies for individual subjects. Thus, while a concentration of 10 ppm insecticide kills the first insect, it may take 12 ppm to kill the next one. The dosage level at which a given subject responds is called the tolerance to the stimulus. Since the tolerance varies from subject to subject, it is necessary to determine the distribution of the tolerances for the population of subjects.

If the distribution of the tolerances is given by f(x), the probability of a response to a dose x is given by

$$P(\text{response} \mid \text{dose}=x) = \int_{-\infty}^{x} f(v)\, dv$$

The adequacy of the dose response model is determined by the suitability of the tolerance distribution assumed. Although Finney lists a number of distributions that have been used, he argues that only the normal and logistic distributions have been given a practical justification. If the logistic distribution is assumed, the method is called Logit Analysis. Probit analysis, the subject of the remainder of this study, is the method that deals with quantal responses where the underlying tolerance distribution is assumed to be a normal.

The problem in dose-response studies is to estimate the dosage levels that will give certain desired response rates. The variable of interest in an experiment will be the number, or percentage, of subjects that exhibited the desired response at each dosage level. Thus the dosage levels are estimated in terms of the dose that causes a given percentage of the subjects to respond. These doses are known as the effective doses or, because of the early methodological contributions of insecticide studies, more commonly as the lethal doses (LD). Thus, LD10 and LD90 are, respectively, the doses which are estimated to cause a 10% and a 90% response rate, and LD50 is the median effective dose (which is the mean for the normal distribution).

In the probit dose response model, where the tolerances are normally distributed, the probability of a response to a dose x is given by:

$$P(\text{response}\,|\,\text{dose}=x) = \int_{-\infty}^{X} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(v-\mu)^2}{2\sigma^2}\right] dv$$

The solution of the LD50 is the dose which causes half of the subjects to respond. For the normal case, as with any symmetric distribution, it is evident that this value is equivalent to the mean of the distribution, that is, LD50=μ. The solution for other LD values will involve both the mean and standard deviation of the tolerance distribution. It is helpful to make the following transformation:

$$P = \int_{-\infty}^{X} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(v-\mu)^2}{2\sigma^2}\right] dv = \int_{-\infty}^{\frac{X-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] dz$$

$$= \int_{-\infty}^{Y} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] dz$$

where $Y = (x - \mu) / \sigma = -\mu/\sigma + x/\sigma = \alpha + \beta x$.

Thus the mean and standard deviation of the model may be expressed in the form of the linear equation $Y = \alpha + \beta x$.

The mean and standard deviation are most commonly estimated by using maximum likelihood estimates. If, in an experiment, a dose x_i is tested on n_i subjects, the probability that r_i respond is the binomial probability given by

$$\binom{n_i}{r_i} P_i^{r_i} (1-P_i)^{n_i-r_i}$$

If k different doses of the stimulus are tested, the log likelihood function, L (say), is proportional to

$$\sum_{i=1}^{k} r_i \ln(P_i) + \sum_{i=1}^{k} (n_i-r_i) \ln(1-P_i)$$

Since

$$P_i = \int_{-\infty}^{Y_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) dz$$

where $Y_i = \alpha + \beta x_i$,

then the maximum likelihood estimates of μ and σ are obtained by determining the values of α and β which maximize the likelihood function. These values are found by solving simultaneously for α and β in

$$\sum_{i=1}^{k} \frac{n_i \left(\frac{r_i}{n_i} - P_i\right) Z_i}{P_i (1-P_i)} = 0$$

and

$$\sum_{i=1}^{k} \frac{n_i \left(\frac{r_i}{n_i} - P_i\right) Z_i X_i}{P_i\,(1-P_i)} = 0$$

where Z_i is the height or ordinate of the standard normal curve at the point Y_i.

Although it may be difficult to solve these equations explicitly, often an iterative procedure will readily provide approximations until a satisfactorily close solution is determined. If, for example, α_1 and β_1 are initial estimates of α and β, second approximations $\alpha_2 = \alpha_1 + \delta\alpha_1$ and $\beta_2 = \beta_1 + \delta\beta_1$ may be obtained by a Taylor-Maclaurin series expansion to the first order, where $\delta\alpha_1$ and $\delta\beta_1$ are the solutions to

$$\frac{\partial L}{\partial \alpha_1} + \delta\alpha_1 \frac{\partial^2 L}{\partial \alpha_1^2} + \delta\beta_1 \frac{\partial^2 L}{\partial \alpha_1 \partial \beta_1} = 0$$

and

$$\frac{\partial L}{\partial \beta_1} + \delta\alpha_1 \frac{\partial^2 L}{\partial \alpha_1 \partial \beta_1} + \delta\beta_1 \frac{\partial^2 L}{\partial \beta_1^2} = 0$$

The second approximations, $\alpha_1 + \delta\alpha_1$ and $\beta_1 + \delta\beta_1$, may then be substituted for the initial estimates α_1 and β_1, and the process may be repeated until a specified measure of closeness is attained.

Finney makes the approximation $r_i/n_i = p_i$ in the second partials and mixed partials above. With this approximation, and letting

$$w_i = Z_i^2 \,/\, p_i(1 - p_i)$$

the equations above simplify to the following equations which have the form of normal equations for weighted least squares

$$\delta\alpha_1 \sum_{i=1}^{k} n_i w_i + \delta\beta_1 \sum_{i=1}^{k} n_i w_i x_i = \sum_{i=1}^{k} n_i w_i \left(\frac{\frac{r_i}{n_i} - P_i}{Z_i} \right)$$

and

$$\delta\alpha_1 \sum_{i=1}^{k} n_i w_i x_i + \delta\beta_1 \sum_{i=1}^{k} n_i w_i x_i^2 = \sum_{i=1}^{k} n_i w_i x_i \left(\frac{\frac{r_i}{n_i} - P_i}{Z_i} \right)$$

where r_i, n_i, x_i, and Z_i are as defined previously. Letting

$$y_i = \alpha_1 + \beta_1 x_i + \frac{\frac{r_i}{n_i} - P_i}{Z_i}, \quad \alpha_2 = \alpha_1 + \delta\alpha_1 \quad \text{and} \quad \beta_2 = \beta_1 + \delta\beta_1$$

then substituting, the normal equations become

$$\alpha_2 \sum_{i=1}^{k} n_i w_i + \beta_2 \sum_{i=1}^{k} n_i w_i x_i = \sum_{i=1}^{k} n_i w_i y_i$$

and

$$\alpha_2 \sum_{i=1}^{k} n_i w_i x_i + \beta_2 \sum_{i=1}^{k} n_i w_i x_i^2 = \sum_{i=1}^{k} n_i w_i x_i y_i$$

It is now evident that solutions to α_2 and β_2 may be obtained by linear regression methods. Letting

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i w_i x_i}{\sum_{i=1}^{k} n_i w_i} \quad \text{and} \quad \bar{y} = \frac{\sum_{i=1}^{k} n_i w_i y_i}{\sum_{i=1}^{k} n_i w_i}$$

the solutions become

$$\beta_2 = \frac{\sum_{i=1}^{k} n_i w_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{k} n_i w_i (x_i - \bar{x})^2}$$

and

$$\alpha_2 = \bar{y} - \beta_2 \bar{x}$$

Now set $\beta_1 = \beta_2$ and $\alpha_1 = \alpha_2$ and iterate for another approximation. The procedure is continued until a prespecified measure of closeness is attained. The usual measure is to compute

$$\sum_{i=1}^{k} (\alpha_1 + \beta_1 x_i - \alpha_2 - \beta_2 x_i)^2$$

after each iteration. When the difference between this sum of squares from successive iterations is small, subsequent iterations will change the estimates of α and β very little for the normal distribution, so the procedure is stopped.

Once the iterative process is completed, a test of the fit of the estimated model to the data may be made. The test statistic, which is a weighted sum of squares of the deviation between the predicted and the observed doses, is given by

$$\chi_c^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

where

$$S_{yy} = \sum_{i=1}^{k} n_i w_i y_i^2 - \frac{\left(\sum_{i=1}^{k} n_i w_i y_i\right)^2}{\sum_{i=1}^{k} n_i w_i}$$

$$S_{xy} = \sum_{i=1}^{k} n_i w_i x_i y_i - \frac{(\sum_{i=1}^{k} n_i w_i x_i)(\sum_{i=1}^{k} n_i w_i y_i)}{\sum_{i=1}^{k} n_i w_i}$$

and

$$S_{xx} = \sum_{i=1}^{k} n_i w_i x_i^2 - \frac{(\sum_{i=1}^{k} n_i w_i x_i)^2}{\sum_{i=1}^{k} n_i w_i}$$

Under the null hypothesis that the model fits the data, X_c^2 is approximately distributed as a chi-square with k-2 degrees of freedom.

If the hypothesis that the model fits is rejected, a further examination of the data will be required. If deviations between the predicted and observed data are systematic, the indication is that the model does not fit, and the researcher must begin with a new model. On the other hand, if the deviations are sporadic about the response curve, the groups of subjects tested at each dose may be heterogeneous. If the researcher believes that the model is correct and that the large deviations are caused by a lack of randomness or independence of the test subjects, Finney defines a 'heterogeneity factor', which is given by

$$h = \frac{X_c^2}{k-2}$$

All variances must be multiplied by this factor, which increases as the deviations from the model increase. Additionally, all uses of the standard

errors, such as in computing confidence intervals, must use a percentage point from the Student's t distribution with k-2 d.f., rather than the normal percentage point. Finney warns against adopting the heterogeneity factor simply for its convenience.

Fiducial limits, which are similar (and often identical) to confidence limits, and which reflect the uncertainty over the true parameter values, may be set about the LD values to provide a range in which the true LD values should lie. The appropriate variance is

$$\mathrm{Var}(Y_i) = \frac{1}{\sum_{i=1}^{k} n_i w_i} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$

The fiducial limits about a given LD value are then

$$Y_i \pm (Z)\ \mathrm{s.e.}(Y_i)$$

where Z is the appropriate percentage point from the standard normal distribution. If a heterogeneity factor is being applied, it must be multiplied against the variance of the LD value and the normal distribution percentage point must be changed to the appropriate percentage point from the Student's t distribution.

A more satisfactory approach to computing the fiducial limits is to use Fieller's theorem for the variance of a ratio. The form of the limits is

$$m + \frac{g}{1-g}(m - \bar{x}) \pm \frac{t}{\beta_1(1-g)} \left(\frac{h(1-g)}{\sum_{i=1}^{k} n_i w_i} + \frac{h(m-\bar{x})^2}{S_{xx}} \right)^{\frac{1}{2}}$$

where: m is the LD estimate (e.g., for the LD50, $m=\alpha_1$),

t is a normal percentage point,

$$g = \frac{t^2 h}{\beta_1^2 S_{xx}}$$

and h = 1.

If the heterogeneity factor is appropriate, it is substituted for h and t becomes the appropriate Student's t percentage point.

III. The Design and Results of a Simulation

The possibility of random variation around dosages in the probit dose-response model was considered in a paper by Patwary and Haley (1967). They indicated several modifications in the probit analysis method that could be used to account for discrete (and specifically, Poisson) variation about the doses. The interest in the current study lies more in determining the seriousness of the effects of dosage errors on specific results of probit analysis. In other words, does the standard probit analysis technique, which assumes that the dosages are fixed, perform satisfactorily even when the dosages are random variables? The present study is limited to the case when the dosage errors are normally distributed.

For the extreme case, it is evident that if variation about the dosages was large enough, the dosages could not be distinguished and an adequate dose-response model could not be fit. Although it may be presumed that no carefully planned study would allow such an extreme, it is also extremely difficult to conceive of a study in which the dosages were exactly fixed with no possibility of variation. The question becomes, then, how much variation is tolerable?

To test the effects of variation about dosages, a simulation study was designed. Three levels of variation were considered. For purposes of control, some groups of subjects were exposed to the 'nominal' dosages with the only variation being the binomial variation of the subject's response.

The other groups of subjects were subjected to 2 different levels of dosage-errors which were simulated to cause the 'actual' dosages to vary about the nominal dosages. For one level of variation, successive levels of the nominal dosages were more than 2 standard deviations of the error apart, while the second variation level caused 2 successive nominal dosage levels to be within 2 standard deviations. The error variation was assumed to be normally distributed with equal variances about each of the nominal dosage levels.

Two other related factors were considered along with the dosage errors. Since the effect of both large and small numbers of subjects was desired, the simulation was constructed for both 10 and 50 subjects per dosage level. The third factor was motivated by a possible difference in the method of application of the dose. In many experiments, a dose or batch of doses given to a group of subjects might vary from the presumed dose, but vary in the same direction for each subject. For example, if a pill was supposed to contain 20 mg of a drug, a certain batch might be slightly shortchanged so that every pill had approximately 19.8 mg of the substance. Or an insecticide sprayed on 50 flies might only have 19.8 ppm concentration of its active ingredient, exposing each subject uniformly to a slightly lower dose. On the other hand, a spray with 20 ppm might be diffused during flight so that some flies received a different concentration of the insecticide than others. The experimental situation may affect how variation in the doses is represented to individual subjects in the experiment. Thus the third factor involved simulating the application of the dosages to 3 different sizes of groups of subjects. In the first case,

all subjects were treated collectively with a uniform dose of the stimulus. The second case divided the subjects into two equal groups which received uniform doses from two separate batches of the stimulus. The third case varied the dosage individually for each subject in the sample.

The resulting 14 treatment combinations of the 3 factors are summarized in the following chart, listing in the appropriate cell the notation with which each treatment combination will be identified in the rest of this chapter (the notation is of the form (N,S,G), where N is the sample size, S is the standard deviation of the dosage errors, and G, if applicable, is the group size to which the doses were applied):

	s=0.0		s=1.0		s=2.0	
	Number of Subjects:		Number of Subjects:		Number of Subjects:	
	10	50	10	50	10	50
Group size:	NA:(10,0)	(50,0)	(10,1,10)	(50,1,50)	(10,2,10)	(50,2,50)
			(10,1,5)	(50,1,25)	(10,2,5)	(50,2,25)
			(10,1,1)	(50,1,1)	(10,2,1)	(50,2,1)

where s is the standard deviation of the dosage errors.

In order to simulate the probit dose-response model, an intercept and slope of the response curve or, equivalently, a mean and standard deviation of the tolerance distribution must be selected. The selection of a value for the mean is arbitrary since the response curve will shift back and forth but will not change in form if different means are simulated. On the other hand, the variance will change the form of the response function since it singularly determines the slope of the function. The selection of a particular value for the variance, however, is important only in relation to

the range of the doses selected. Since the performance of probit analysis is primarily affected by the range of the response rate, Finney (1971) has recommended that an experiment should aim for doses that give response rates of between 10% and 90% for small sample sizes and between 30% and 70% for large samples.

For this study, a dose of 20.0 was assumed to be the median effective dose. A standard deviation of 6.0 was assumed and 5 doses were selected at the levels 15.0, 17.5, 20.0, 22.5, and 25.0. Although these values are arbitrary, this combination of dosage levels and variance will keep the response rate roughly in the range of 10-90% (even with some dosage variation).

The data were simulated independently for each of the 14 experimental conditions making the design of the simulation study a completely randomized design. For the nominal case, the expected probability of response for each dose was computed by integrating the normal distribution up to the given dosage level. Similarly, a probability was computed for the dosage-error cases after a randomly generated normal variation was added to each dosage level. The normal random numbers were generated using the Statistical Analysis System (SAS) Institute's RANNOR function. Once the probabilities were calculated, they were compared in magnitude to a stream of either 10 or 50 uniform random numbers, corresponding to subjects, using the SAS RANUNI function. If P is the probability of response for a given dose and U represents the uniform variable, then if $U < P$, the subject was recorded as responding to the dose, while if $U > P$ a non-response was recorded. The procedure was then replicated for each of the 14 treatment combinations, 200

times for the 2 nominal dose combinations and 50 times for the 12 dosage-error combinations. The random number streams used for each of the 10 treatment combinations were seeded independently.

The effects of the 3 factors in the study on the estimates of the population parameters were investigated. The population parameters are μ, σ, and LDp, where μ is the model mean, σ is the model standard deviation, and LDp are the LD values for percentages from .01 to .99. Means and variances of the estimates were obtained from the simulation. In addition, fiducial limits of the probit dose response model were considered. For each percentile for which an LDp value was computed, it was observed whether the expected dose at that percentile fell within or without the estimated 95% fiducial limits.

Results of the Simulation

The maximum likelihood estimates of the mean and variance of the probit model, as described by Finney, and the estimated 95% fiducial limits have been incorporated into the SAS Institute's Probit Analysis procedure, which was used to compute these estimates. The pertinent estimates of the mean and standard deviation, with their sample variances and covariance, are summarized in Table 1 for all 14 treatment combinations.

In general it may be seen that each of the 14 treatment combinations closely estimate the true value of the mean. The population mean, 20, is within a standard deviation of each treatment's estimate of the mean. Thus,

the estimate of the mean of the probit model appears to be unbiased over the levels of the factors considered in this study.

Table 1. Estimated means and standard deviations of the model, with their sample variance and covariance.

Treatment	$\hat{\mu}$	$Var(\hat{\mu})$	$\hat{\sigma}$	$Var(\hat{\sigma})$	$Cov(\hat{\mu},\hat{\sigma})$
(10,0)	20.1342	3.8544	6.99332	48.557	9.182
(10,1,1)	20.0363	2.1538	7.04446	15.394	1.526
(10,2,1)	20.0337	1.3132	6.59288	6.388	0.465
(10,1,5)	20.3736	2.7500	7.09462	20.998	4.713
(10,2,5)	20.1145	2.0510	6.93068	20.866	2.918
(10,1,10)	20.2781	2.8473	8.52665	179.894	8.805
(10,2,10)	19.7968	20.1403	8.72122	174.458	-38.998
(50,0)	20.0077	0.2452	6.05314	0.749	-0.071
(50,1,1)	19.8599	0.2685	6.24935	1.000	0.043
(50,2,1)	20.0882	0.3087	6.39461	1.641	0.144
(50,1,25)	20.2050	0.2288	6.05205	0.743	0.194
(50,2,25)	20.0175	0.6283	6.95123	5.467	-0.639
(50,1,50)	20.0925	0.3100	5.95700	0.890	0.087
(50,2,50)	19.9957	1.8991	6.56063	3.401	-0.636

The sample variances of the estimated means for the 14 treatments, which are shown in column 2 of Table 1, indicate a different picture of the effect of the experimental factors on the estimation of the means. The magnitude of the higher variances of the smaller sample size treatments are particularly noteworthy. While the variances of $\hat{\mu}$ are generally below 1.0 for the large sample size treatments, treatment (10,2,10), with a small sample size and large dosage errors, has a sample variance of over 20.

The variances of the estimated means for the 14 treatments were tested for equality using Levene's test. For each treatment, the absolute value of the deviation of each observed model's estimate of the mean from the

treatment sample mean was recorded. An analysis of variance conducted on these absolute deviations gave an approximate test of the equality of variances from each treatment. The main factor effects are tested by comparing the means of the deviations over each given level of the factors. The analysis of variance table, followed by an ordered listing of pairwise t-tests on the 14 treatment mean deviations, is given in Table 2. The results indicate that there are significant interactions between the level of dosage error and the method of application, and between dosage error and sample size. The unusually high variance of treatment (10,2,10) indicates that when a small sample size is combined with large dosage errors and a collective dose application, the estimate of the mean of the probit model lacks precision. Most of the small sample size treatments may be significantly differentiated from the large sample size treatments. The larger dosage error treatments demonstrate less precision of the estimate when the dose is applied collectively, but an application of the doses to the subjects individually or to subgroups of the subjects negates to some extent the dosage error effect.

The difference in the precision of the estimates of the mean by different treatment combinations is further illustrated by the boxplots of the sample means given for each treatment in Figure 1. The plots demonstrate that the distributions of the small sample estimates are subject to more extreme values than the larger sample size treatments. Thus, although the estimates of the mean by all treatment combinations are very good on the average, the small sample cases may give very poor estimates for any given trial. Table 3 shows the results of a test of normality of the distribution of each treatment: the Shapiro-Wilk test was used for the nominal treatments, and the Kolmogorov-Smirnov test for the other

Table 2. Levene's test on the equality of variances of the estimated model means.

Source	DF	SUM OF SQUARES	F	PR > F
Sigma	2	17.03333521	5.40	0.0047
N	1	103.99698409	65.89	0.0001
Group size	2	31.84284328	10.09	0.0001
S*N	2	27.44238609	8.69	0.0002
S*G	2*	28.33466490	8.98	0.0001
N*G	2	4.48708993	1.42	0.2418
S*N*G	2*	3.91005949	1.24	0.2902
Error	986	1556.21378202		
Total	999	1780.04569667		

*Note: within levels of G, there are only 2 levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	2.2718	50	(10,2,10)
B	1.1841	50	(10,1,10)
B	1.1826	50	(10,1,5)
B	1.1604	200	(10,0)
B	1.1051	50	(50,2,50)
B	1.1034	50	(10,1,1)
B C	1.0574	50	(10,2,5)
B C D	0.9096	50	(10,2,1)
C D E	0.6097	50	(50,2,25)
D E	0.4371	50	(50,1,1)
D E	0.4236	50	(50,1,50)
D E	0.4230	50	(50,2,1)
E	0.3881	50	(50,1,25)
E	0.3837	200	(50,0)

Figure 1. Boxplots of the estimated model means for each of the 14 treatments.

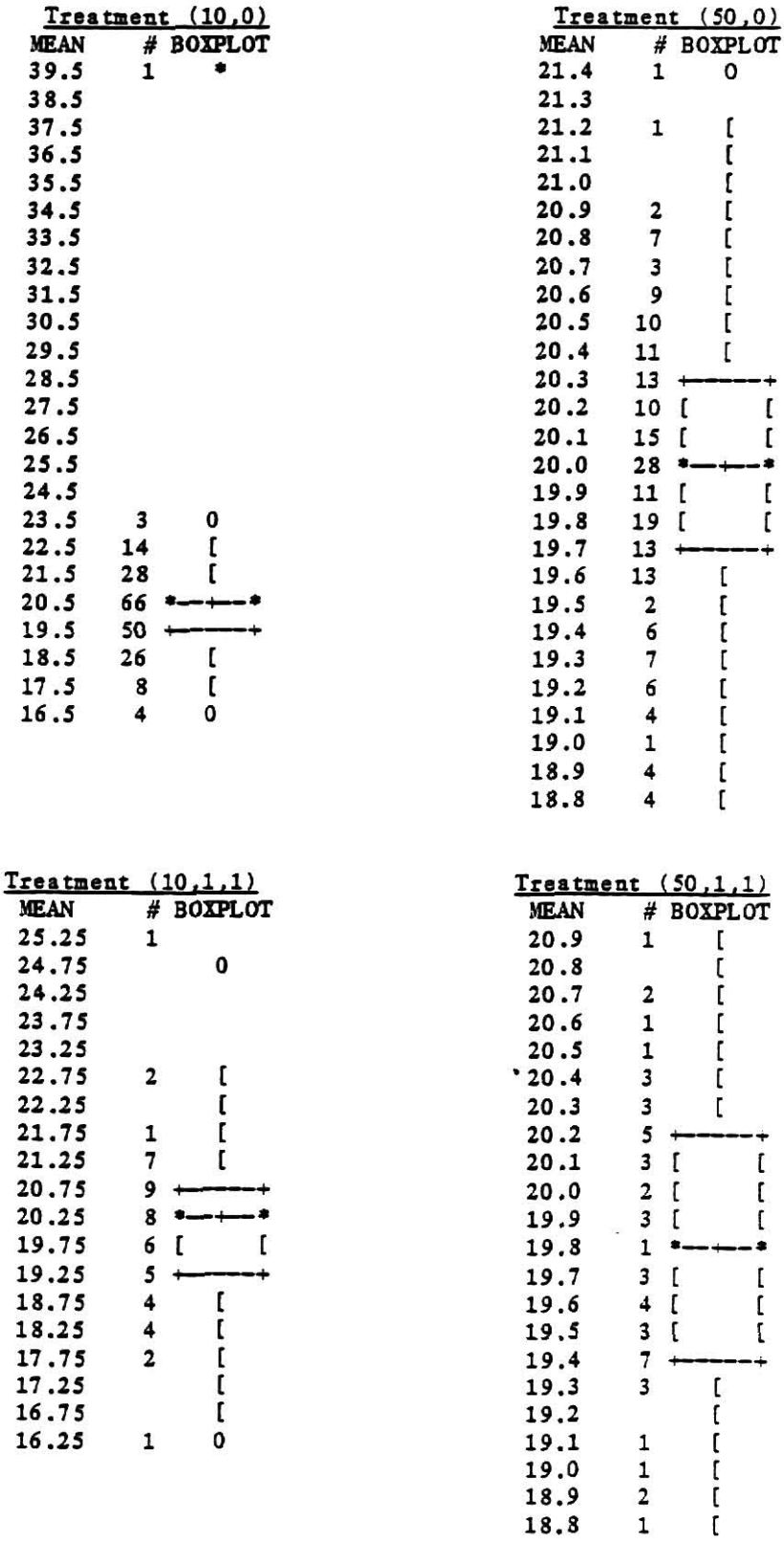


Figure 1. continued

Treatment (10,2,1)

MEAN	#	BOXPLOT
22.75	2	[
22.25	1	[
21.75	2	[
21.25	5	[
20.75	8	+-----+
20.25	10	[+ [
19.75	7	*-----*
19.25	6	+-----+
18.75	6	[
18.25	1	[
17.75	2	[

Treatment (50,2,1)

MEAN	#	BOXPLOT
21.4	1	0
21.2		[
21.0	1	[
20.8	2	[
20.6	5	[
20.4	4	+-----+
20.2	9	[[
20.0	7	*--+--*
19.8	6	[[
19.6	8	+-----+
19.4	3	[
19.2	1	[
19.0	1	[
18.8	1	[
18.6		
18.4	1	0

Treatment (10,1,5)

Mean	#	BOXPLOT
26.75	1	*
26.25		
25.75		
25.25		
24.75	1	0
24.25		
23.75		
23.25		
22.75	4	[
22.25		[
21.75	2	[
21.25	7	+-----+
20.75	5	[[
20.25	6	*--+--*
19.75	11	[[
19.25	4	+-----+
18.75	3	[
18.25	4	[
17.75	2	[

Treatment (50,1,25)

Mean	#	BOXPLOT
21.0	1	[
20.9	3	[
20.8	4	[
20.7	1	[
20.6	3	[
20.5	3	+-----+
20.4	2	[[
20.3	5	[[
20.2	4	*--+--*
20.1	2	[[
20.0	6	[[
19.9	3	[[
19.8	5	+-----+
19.7	1	[
19.6	2	[
19.5		[
19.4	2	[
19.3	2	[
19.2		[
19.1	1	[

Figure 1. continued.

Treatment (10,2,5)

Mean	#	BOXPLOT
25.25	1	0
24.75		
24.25		
23.75		
23.25		
22.75	1	[
22.25	1	[
21.75	4	[
21.25	6	+-----+
20.75	4	[[
20.25	9	*--+--*
19.75	10	[[
19.25	3	+-----+
18.75	6	[
18.25	4	[
17.75		[
17.25		[
16.75	1	[

Treatment (50,2,25)

Mean	#	BOXPLOT
21.4	1	[
21.2		[
21.0		[
20.8	5	[
20.6	9	+-----+
20.4	5	[[
20.2	5	[[
20.0	1	*--+--*
19.8	5	[[
19.6	10	+-----+
19.4	1	[
19.2	2	[
19.0	1	[
18.8	2	[
18.6		[
18.4	1	[
18.2		
18.0	1	0
17.8		
17.6		
17.4		
17.2	1	0

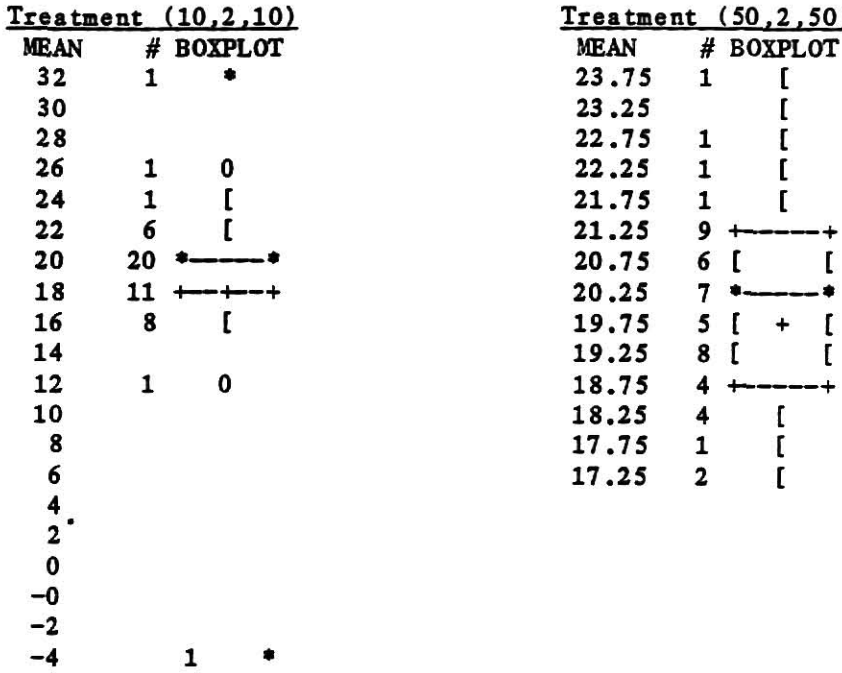
Treatment (10,1,10)

MEAN	#	BOXPLOT
25.25	1	
24.75	1	0
24.25		
23.75		
23.25	1	[
22.75		[
22.25	3	[
21.75	4	[
21.25	4	+-----+
20.75	7	[[
20.25	9	*--+--*
19.75	6	[[
19.25	5	+-----+
18.75	6	[
18.25	1	[
17.75		[
17.25		[
16.75	1	[
16.25		0
15.75		
15.25	1	0

Treatment (50,1,50)

MEAN	#	BOXPLOT
21.2	2	[
21.0		[
20.8	2	[
20.6	6	[
20.4	4	+-----+
20.2	7	[[
20.0	9	*--+--*
19.8	6	[[
19.6	6	+-----+
19.4	3	[
19.2	2	[
19.0	1	[
18.8		[
18.6	2	[

Figure 1. continued



treatments. These tests demonstrate a lack of normality of 5 separate treatment combinations, of which all but one have the small sample size and one or more extreme estimates. These attributes of the small sample size treatment estimates hurt the precision of the estimates.

Table 3. Test of normality of the estimated model means.

Treatment	Test	statistic	p-value	Treatment	Test	statistic	p-value
(10,0)	D:	.117737	.01	(50,0)	D:	.055290	.138
(10,1,1)	W:	.977309	.620	(50,1,1)	W:	.973365	.486
(10,2,1)	W:	.985683	.911	(50,2,1)	W:	.988021	.950
(10,1,5)	W:	.907292	.01	(50,1,25)	W:	.970651	.429
(10,2,5)	W:	.976718	.598	(50,2,25)	W:	.919436	.01
(10,1,10)	W:	.951692	.079	(50,1,50)	W:	.976475	.589
(10,2,10)	W:	.703889	.01	(50,2,50)	W:	.985725	.912

To this point, it is evident that the levels of factors considered in this study do not on the average affect the estimate of the mean of the probit model, but the precision of this estimate is affected. In individual trials, a small sample size will occasionally lead to very poor estimates, but these are balanced out if a number of trials are conducted. A smaller sample size also generally hurts the precision of the estimate. The larger level of dosage errors is typically detrimental, especially when the doses are applied collectively, but an individual application will sometimes balance out the dosage error effects.

The estimate of the standard deviation of the probit model for each of the 14 treatments has been given in Table 1. Each treatment provides a reasonable estimate of the standard deviation, as the true population standard deviation of 6.0 is within one sample standard deviation of each treatment's estimate. At the same time, however, the small sample size treatments consistently overestimate the variance. For each pair of

corresponding small-large sample size treatments except one, the small sample size treatment estimate is farther away from the population value than is the large sample size estimate. This difference is particularly noticeable in the collective application treatments.

The variances of the estimated model standard deviations magnify the effect that a small sample size has on the estimates. The extremely high variances of some of the small sample size treatments was unexpected. Treatments (10,1,10) and (10,2,10), with small sample sizes and a collective application of the doses, have variances which are over 50 times as great as the variances of the corresponding large sample treatments.

Levene's test was used to test the equality of the variances of the estimated standard deviations for the 14 treatments. The analysis of variance table of the absolute differences of the standard deviations from each treatment's average standard deviation is given in Table 4, followed by an ordered listing of pairwise t-tests between the 14 treatments. The interaction between sample size and application method is significant because of the detrimental effect of a small sample size when a collective application is used but not when the dosages are applied to smaller groups of subjects. Similarly, the interaction between sample size and level of dosage errors is caused by a detrimental effect of the dosage errors on the small sample size treatments but the lack of a dosage error effect for large sample size treatments. Thus the smaller sample size will generally hurt the precision of the estimate of the standard deviation of the probit model, and especially when the dosages have been applied collectively. The presence of dosage errors also causes a slightly detrimental effect on the precision of the estimates, although the volatibility of the small sample sizes may mask this effect.

Table 4. Levene's test of the equality of variance of the estimated model standard deviations.

Source	DF	SUM OF SQUARES	F	PR > F
Sigma	2	102.75464673	1.98	0.1380
N	1	971.14042179	37.52	0.0001
Group size	2	262.29932698	5.07	0.0065
S*N	2	401.95960170	7.76	0.0005
S*G	2*	28.05792181	0.54	0.5818
N*G	2	219.40457932	4.24	0.0147
S*N*G	2*	7.05957100	0.14	0.8725
Error	986	25523.05416722		
Total	999	27258.22365852		

*Note: within levels of G, there are only 2 levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	5.3979	50	(10,2,10)
A	4.7886	50	(10,1,10)
B	2.7425	50	(10,1,5)
B C	2.6472	50	(10,1,1)
B C	2.5919	50	(10,2,5)
B C	2.4029	200	(10,0)
B C D	1.7905	50	(10,2,1)
B C D	1.5358	50	(50,2,25)
B C D	1.4098	50	(50,2,50)
B C D	0.9069	50	(50,2,1)
B C D	0.8298	50	(50,1,1)
B C D	0.7478	50	(50,1,50)
C D	0.6783	50	(50,1,25)
D	0.6717	200	(50,0)

The boxplots of the sample standard deviations for each treatment, which are given in Figure 2, further evidence the problems caused by small sample sizes. Each small sample size treatment provided at least one greatly exaggerated estimate of the variance. Even without these extreme values, the dispersal of the estimates was greater for the small sample size cases, indicating a general problem with the precision of the estimates.

One might expect that the estimates of the mean and variance of the probit model are independent, but the sample data give conflicting evidence on this point. Table 5 shows the sample correlation between these estimates and a p-value for the test that the sample correlation equals zero. In general it may be seen that the estimates of the model's mean and variance are uncorrelated for the large sample size treatments and correlated for the small sample size treatments. Treatments (50,0), (50,1,25), and (50,2,25) break this pattern, although the estimates that the first two of these treatments give for both the mean and variance are very precise.

Table 5. Correlation coefficients of the estimated mean and variance of the probit model.

Treatment	Corr($\hat{\mu}$,$\hat{\sigma}$)	p-value	Treatment	Corr($\hat{\mu}$,$\hat{\sigma}$)	p-value
(10,0)	.67115	.0001	(50,0)	-.16594	.0189
(10,1,1)	.26503	.0629	(50,1,1)	.08319	.5657
(10,2,1)	.16051	.2655	(50,2,1)	.20253	.1584
(10,1,5)	.62016	.0001	(50,1,25)	.47134	.0006
(10,2,5)	.44616	.0012	(50,2,25)	-.34495	.0142
(10,1,10)	.38904	.0052	(50,1,50)	.16647	.2479
(10,2,10)	-.65791	.0001	(50,2,50)	-.25027	.0796

p-value: for the test of Ho: ρ=0

The estimate of an LDp value is given by $\widehat{LDp} = \hat{\mu} + Z_p\hat{\sigma}$. Estimates of the expected values of the LD estimators for several response probabilities from .01 to .99 are given in Table 6 for all 14 treatment combinations. The

Figure 2. Boxplots of the estimated model standard deviations for each of the 14 treatments.

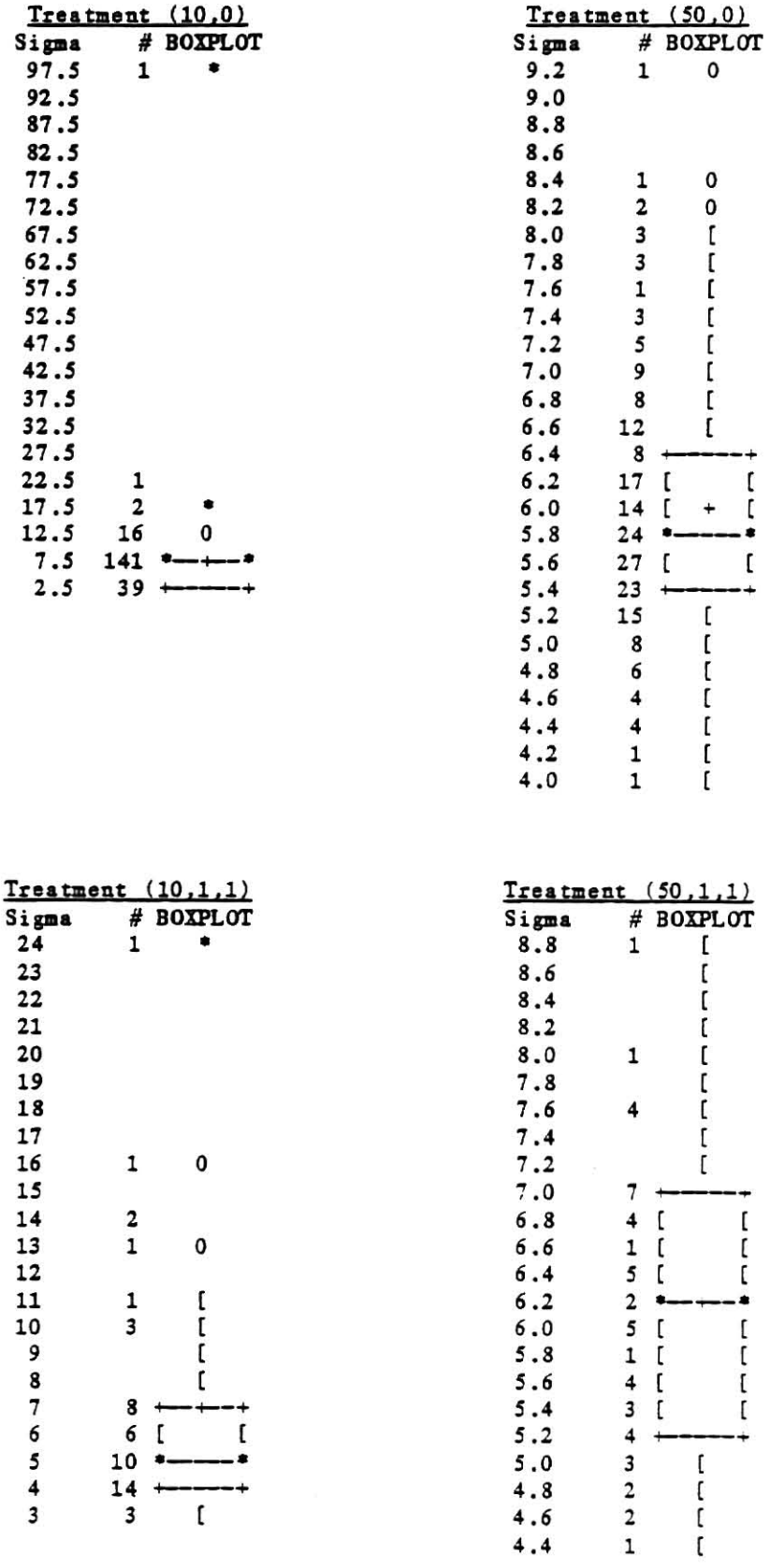


Figure 2. continued

Treatment (10,2,1)

Sigma	#	BOXPLOT
16	1	*
15		
14		
13		
12	1	0
11	1	0
10	1	[
9	2	[
8	4	[
7	9	+-----+
6	10	*--+--*
5	9	+-----+
4	5	[
3	7	[

Treatment (50,2,1)

Sigma	#	BOXPLOT
11.75	1	*
11.25		
10.75		
10.25		
9.75		
9.25		
8.75	3	[
8.25		[
7.75	1	[
7.25	9	+-----+
6.75	7	[[
6.25	10	*--+--*
5.75	9	[[
5.25	7	+-----+
4.75	2	[
4.25	1	[

Treatment (10,1,5)

Sigma	#	BOXPLOT
32	1	*
30		
28		
26		
24		
22		
20		
18		
16		
14		
12	2	[
10	4	[
8	6	+-----+
6	11	[+ [
4	20	*-----*
2	6	[

Treatment (50,1,25)

Sigma	#	BOXPLOT
8.2	1	0
8.0		[
7.8	1	[
7.6		[
7.4	3	[
7.2		[
7.0	3	[
6.8	1	[
6.6	4	+-----+
6.4	2	[[
6.2	4	[[
6.0	7	*--+--*
5.8	4	[[
5.6	4	[[
5.4	5	+-----+
5.2	2	[
5.0	4	[
4.8	3	[
4.6	1	[
4.4	1	[

Figure 2. continued.

Treatment (10,2,5)

Sigma	#	BOXPLOT
32	1	*
30		
28		
26		
24		
22		
20		
18		
16		
14		
12	3	0
10	2	[
8	5	[
6	15	*—+—*
4	18	+-----+
2	6	[

Treatment (50,2,25)

Sigma	#	BOXPLOT
18	1	*
17		
16		
15		
14		
13	1	0
12		
11		
10	1	[
9	2	[
8	5	[
7	10	+-----+
6	11	*—+—*
5	15	+-----+
4	3	[
3	1	[

Treatment (10,1,10)

Sigma	#	BOXPLOT
97.5	1	*
92.5		
87.5		
82.5		
77.5		
72.5		
67.5		
62.5		
57.5		
52.5		
47.5		
42.5		
37.5		
32.5		
27.5		
22.5		
17.5	1	0
12.5	5	[
7.5	32	*—+—*
2.5	11	+-----+

Treatment (50,1,50)

Sigma	#	BOXPLOT
9.25	1	
8.75		0
8.25		
7.75	3	[
7.25	3	[
6.75	6	[
6.25	11	+-----+
5.75	10	*—+—*
5.25	10	+-----+
4.75	5	[
4.25	1	[
3.75		

Figure 2. continued

Treatment (10,2,10)

Sigma	#	BOXPLOT
97.5	1	*
92.5		
87.5		
82.5		
77.5		
72.5		
67.5		
62.5		
57.5		
52.5		
47.5		
42.5		
37.5		
32.5		
27.5		
22.5	2	*
17.5	1	0
12.5	5	0
7.5	26	*—+—*
2.5	15	+-----+

Treatment (50,2,50)

Sigma	#	BOXPLOT
11.75	1	0
11.25		
10.75	1	[
10.25		[
9.75	1	[
9.25	2	[
8.75	4	[
8.25	2	[
7.75	1	[
7.25	3	+-----+
6.75	10	[+ [
6.25	6	*-----*
5.75	5	[[
5.25	1	[[
4.75	10	+-----+
4.25		[
3.75	3	[

Table 6. Mean of the LD values for each response probability.

Prob.	Pop'n	(10,0)	(10,1,1)	(10,2,1)	(10,1,5)	(10,2,5)	(10,1,10)	(10,2,10)
0.01	6.0419	3.8653	3.6484	4.6964	3.8691	3.9914	0.4421	-0.4918
0.02	7.6775	5.7717	5.5688	6.4936	5.8031	5.8807	2.7665	1.8856
0.03	8.7152	6.9812	6.7871	7.6339	7.0301	7.0794	4.2412	3.3940
0.04	9.4959	7.8911	7.7037	8.4917	7.9532	7.9811	5.3506	4.5287
0.05	10.1309	8.6313	8.4492	9.1894	8.7040	8.7146	6.2530	5.4517
0.06	10.6714	9.2612	9.0838	9.7833	9.3431	9.3389	7.0211	6.2373
0.07	11.1453	9.8136	9.6402	10.3040	9.9034	9.8863	7.6945	6.9261
0.08	11.5696	10.3081	10.1383	10.7702	10.4052	10.3764	8.2975	7.5429
0.09	11.9555	10.7579	10.5914	11.1943	10.8615	10.8222	8.8459	8.1038
0.10	12.3107	11.1719	11.0085	11.5846	11.2815	11.2325	9.3508	8.6201
0.15	13.7814	12.8861	12.7352	13.2006	13.0205	12.9314	11.4408	10.7579
0.20	14.9503	14.2485	14.1075	14.4850	14.4026	14.2815	13.1019	12.4569
0.25	15.9531	15.4173	15.2849	15.5869	15.5884	15.4399	14.5270	13.9145
0.30	16.8536	16.4669	16.3422	16.5764	16.6532	16.4801	15.8067	15.2234
0.35	17.6881	17.4396	17.3219	17.4933	17.6399	17.4440	16.9926	16.4364
0.40	18.4799	18.3625	18.2516	18.3634	18.5762	18.3587	18.1179	17.5873
0.45	19.2460	19.2554	19.1511	19.2052	19.4821	19.2436	19.2066	18.7009
0.50	20.0000	20.1342	20.0363	20.0337	20.3736	20.1145	20.2781	19.7968
0.55	20.7540	21.0130	20.9215	20.8622	21.2651	20.9855	21.3496	20.8928
0.60	21.5201	21.9060	21.8210	21.7040	22.1710	21.8704	22.4383	22.0063
0.65	22.3119	22.8289	22.7507	22.5741	23.1073	22.7851	23.5636	23.1573
0.70	23.1464	23.8015	23.7304	23.4910	24.0940	23.7490	24.7495	24.3702
0.75	24.0469	24.8512	24.7877	24.4805	25.1589	24.7892	26.0292	25.6792
0.80	25.0497	26.0200	25.9651	25.5824	26.3446	25.9476	27.4543	27.1368
0.85	26.2186	27.3823	27.3374	26.8668	27.7267	27.2977	29.1154	28.8358
0.90	27.6893	29.0965	29.0642	28.4828	29.4657	28.9966	31.2054	30.9735
0.91	28.0445	29.5106	29.4812	28.8732	29.8858	29.4069	31.7102	31.4899
0.92	28.4304	29.9604	29.9343	29.2972	30.3421	29.8526	32.2586	32.0508
0.93	28.8547	30.4549	30.4325	29.7634	30.8438	30.3428	32.8617	32.6675
0.94	29.3286	31.0073	30.9889	30.2841	31.4041	30.8902	33.5351	33.3564
0.95	29.8691	31.6372	31.6234	30.8780	32.0432	31.5145	34.3032	34.1420
0.96	30.5041	32.3773	32.3690	31.5758	32.7941	32.2480	35.2056	35.0650
0.97	31.2848	33.2872	33.2855	32.4336	33.7171	33.1497	36.3150	36.1997
0.98	32.3225	34.4968	34.5039	33.5738	34.9442	34.3484	37.7897	37.7080
0.99	33.9581	36.4031	36.4242	35.3710	36.8782	36.2377	40.1141	40.0854

Table 6. (continued)

Prob.	Pop'n	(50,0)	(50,1,1)	(50,2,1)	(50,1,25	(50,2,25)	(50,1,50)	(50,2,50)
0.01	6.0419	5.9260	5.3218	5.2121	6.1258	3.8465	6.2344	4.7334
0.02	7.6775	7.5761	7.0253	6.9553	7.7756	5.7414	7.8583	6.5218
0.03	8.7152	8.6230	8.1062	8.0613	8.8223	6.9437	8.8886	7.6565
0.04	9.4959	9.4106	8.9193	8.8933	9.6097	7.8481	9.6636	8.5101
0.05	10.1309	10.0512	9.5807	9.5700	10.2502	8.5837	10.2941	9.2045
0.06	10.6714	10.5965	10.1436	10.1461	10.7954	9.2099	10.8307	9.7954
0.07	11.1453	11.0746	10.6372	10.6511	11.2734	9.7589	11.3012	10.3136
0.08	11.5696	11.5027	11.0792	11.1033	11.7014	10.2505	11.7225	10.7776
0.09	11.9555	11.8920	11.4811	11.5146	12.0907	10.6976	12.1056	11.1995
0.10	12.3107	12.2503	11.8511	11.8932	12.4490	11.1091	12.4583	11.5879
0.15	13.7814	13.7341	13.3829	13.4606	13.9324	12.8130	13.9185	13.1961
0.20	14.9503	14.9133	14.6004	14.7064	15.1114	14.1672	15.0789	14.4742
0.25	15.9531	15.9250	15.6448	15.7751	16.1229	15.3290	16.0746	15.5707
0.30	16.8536	16.8335	16.5828	16.7349	17.0313	16.3723	16.9686	16.5553
0.35	17.6881	17.6754	17.4519	17.6243	17.8730	17.3390	17.7971	17.4678
0.40	18.4799	18.4742	18.2767	18.4682	18.6717	18.2564	18.5833	18.3336
0.45	19.2460	19.2471	19.0746	19.2847	19.4445	19.1440	19.3439	19.1713
0.50	20.0000	20.0077	19.8599	20.0882	20.2050	20.0175	20.0925	19.9957
0.55	20.7540	20.7684	20.6452	20.8918	20.9655	20.8910	20.8411	20.8202
0.60	21.5201	21.5413	21.4432	21.7083	21.7382	21.7786	21.6017	21.6579
0.65	22.3119	22.3401	22.2679	22.5522	22.5370	22.6959	22.3878	22.5237
0.70	23.1464	23.1820	23.1371	23.4416	23.3787	23.6627	23.2163	23.4361
0.75	24.0469	24.0905	24.0751	24.4013	24.2870	24.7060	24.1104	24.4208
0.80	25.0497	25.1022	25.1195	25.4701	25.2985	25.8678	25.1060	25.5173
0.85	26.2186	26.2814	26.3370	26.7158	26.4775	27.2220	26.2665	26.7954
0.90	27.6893	27.7652	27.8688	28.2833	27.9610	28.9258	27.7267	28.4035
0.91	28.0445	28.1235	28.2388	28.6618	28.3193	29.3374	28.0794	28.7919
0.92	28.4304	28.5128	28.6407	29.0731	28.7085	29.7845	28.4625	29.2139
0.93	28.8547	28.9409	29.0827	29.5253	29.1365	30.2760	28.8838	29.6779
0.94	29.3286	29.4190	29.5763	30.0304	29.6145	30.8251	29.3543	30.1960
0.95	29.8691	29.9643	30.1392	30.6064	30.1597	31.4512	29.8909	30.7870
0.96	30.5041	30.6049	30.8006	31.2832	30.8002	32.1869	30.5213	31.4813
0.97	31.2848	31.3925	31.6137	32.1152	31.5876	33.0913	31.2964	32.3349
0.98	32.3225	32.4394	32.6945	33.2212	32.6344	34.2936	32.3267	33.4696
0.99	33.9581	34.0895	34.3981	34.9643	34.2841	36.1885	33.9505	35.2580

estimates were obtained by averaging over all replications in the simulation. Column 1 of Table 6 gives the true LD values for the population. The estimates are plotted in Figure 3 for the small sample size treatments and in Figure 4 for the large sample size treatments.

The LD50 parameter is identical to the mean of the probit model. It has already been indicated that the levels of the factors considered in this study do not bias the estimate of the model's mean. In general it may be seen that the other estimated LD values in the middle percentiles are very close to the population values for this model. In the low and high percentiles, however, as the model's variance has an increased influence on the LD values, the estimated LD values are much farther away from the population values than in the middle percentiles. This difference in the extreme percentiles reflects the exaggerated estimate of the model's variance by some of the treatments.

Although the estimates of the LD parameters are not as accurate in the low and high percentiles, all of the estimates are within a standard deviation of the population values. Thus the levels of experimental factors considered do not significantly affect the bias of the LD estimates. The treatments with small sample sizes, high dosage errors, and a collective dosage application show increased, but still nonsignificant, differences between the LD estimates and the population values.

Estimates of the variances of the LD estimators are given in Table 7, and plotted for the small sample size treatments in Figure 5 and for the large sample treatments in Figure 6. The estimates were obtained empirically from the replications, which is equivalent to estimating $\mathrm{Var}(\hat{\mu} + Z\hat{\sigma}) = \mathrm{Var}(\hat{\mu}) + Z^2\mathrm{Var}(\hat{\sigma}) + 2Z\mathrm{Cov}(\hat{\mu},\hat{\sigma})$. The sample variances are

Figure 3. Sample mean of the estimated LD values for the small sample size treatments.

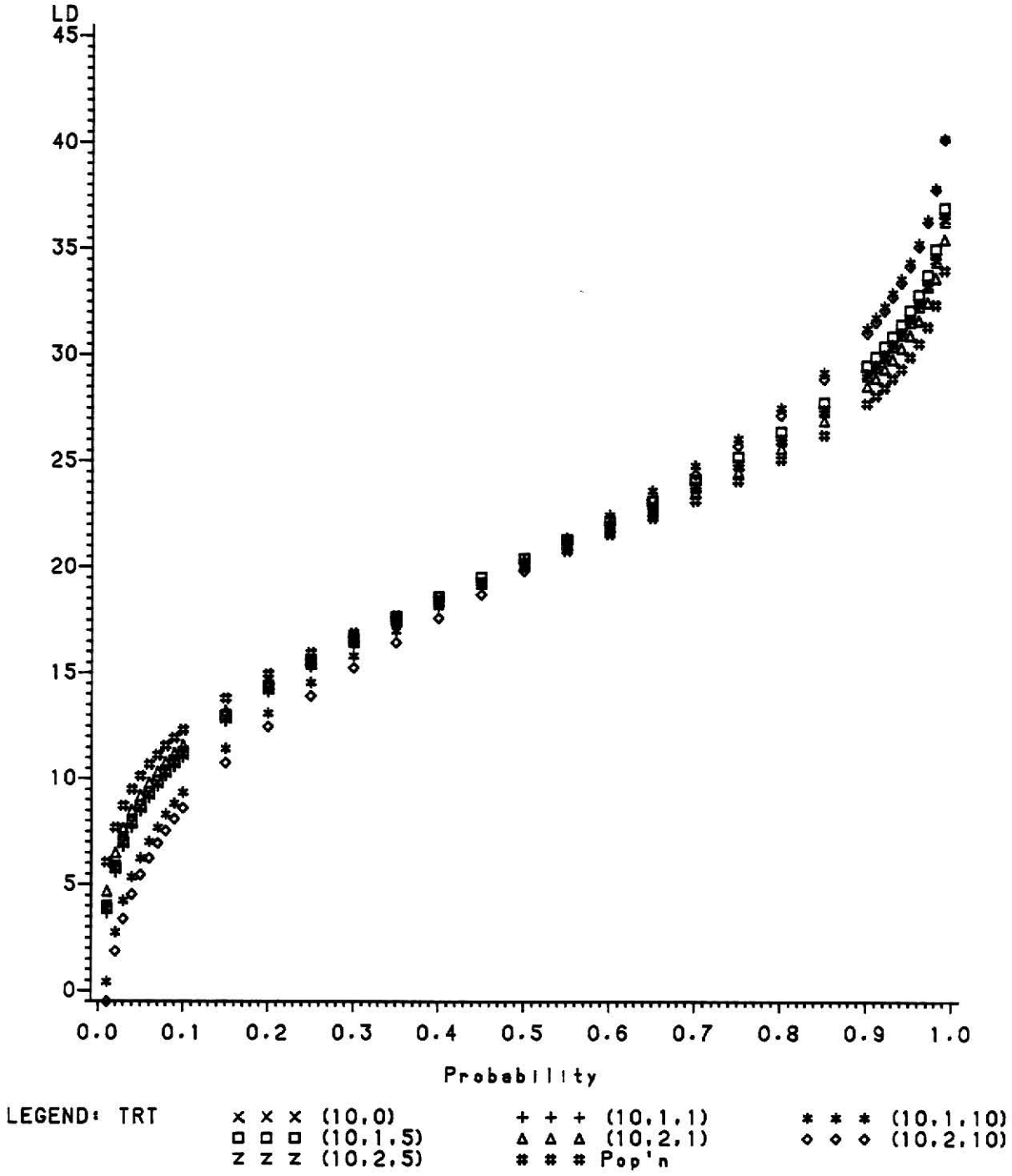


Figure 4. Sample mean of the estimated LD values for the large sample size treatments.

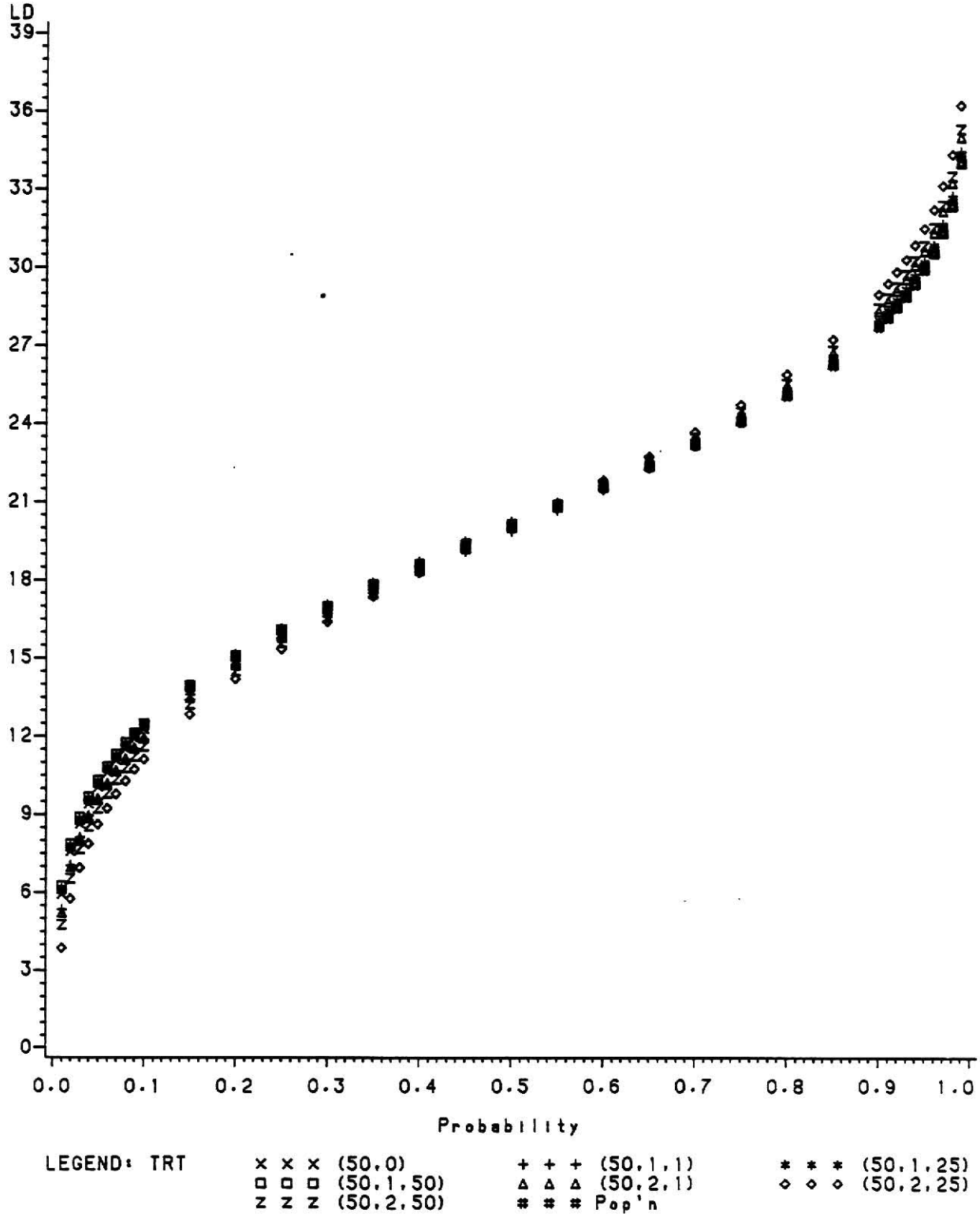


Table 7. Sample variance of the LD values for each response probability.

Prob.	(10,0)	(10,1,1)	(10,2,1)	(10,1,5)	(10,2,5)	(10,1,10)	(10,2,10)
0.01	223.92	78.36	33.72	94.46	101.40	935.45	1145.74
0.02	170.95	60.82	26.35	71.96	78.07	725.45	916.17
0.03	141.08	50.87	22.16	59.30	64.88	606.08	783.96
0.04	120.53	43.99	19.27	50.61	55.79	523.38	691.38
0.05	105.02	38.78	17.07	44.06	48.90	460.59	620.44
0.06	92.68	34.62	15.31	38.86	43.42	410.33	563.13
0.07	82.51	31.18	13.85	34.57	38.88	368.66	515.21
0.08	73.92	28.26	12.62	30.96	35.04	333.26	474.15
0.09	66.52	25.73	11.55	27.86	31.73	302.62	438.32
0.10	60.07	23.52	10.61	25.16	28.84	275.73	406.62
0.15	36.98	15.53	7.21	15.54	18.42	177.84	288.38
0.20	22.79	10.49	5.06	9.69	11.92	115.45	209.36
0.25	13.56	7.10	3.59	5.95	7.61	72.81	152.12
0.30	7.58	4.79	2.58	3.58	4.73	43.08	109.02
0.35	3.99	3.26	1.90	2.24	2.90	22.77	76.10
0.40	2.32	2.37	1.49	1.71	1.91	9.93	51.10
0.45	2.31	2.01	1.30	1.90	1.65	3.48	32.70
0.50	3.85	2.15	1.31	2.75	2.05	2.85	20.14
0.55	6.93	2.78	1.53	4.27	3.11	7.90	13.09
0.60	11.62	3.92	1.96	6.49	4.87	18.85	11.58
0.65	18.14	5.62	2.62	9.50	7.40	36.34	15.99
0.70	26.84	7.99	3.56	13.47	10.85	61.55	27.21
0.75	38.33	11.22	4.85	18.66	15.48	96.56	46.90
0.80	53.70	15.63	6.62	25.56	21.74	145.09	78.07
0.85	75.05	21.85	9.14	35.08	30.51	214.34	126.70
0.90	107.14	31.35	13.00	49.32	43.80	320.87	206.71
0.91	115.76	33.92	14.04	53.13	47.39	349.84	229.18
0.92	125.52	36.83	15.23	57.45	51.45	382.74	254.97
0.93	136.71	40.19	16.60	62.39	56.11	420.64	285.00
0.94	149.78	44.11	18.20	68.16	61.57	465.09	320.60
0.95	165.43	48.82	20.13	75.06	68.11	518.52	363.85
0.96	184.83	54.68	22.52	83.61	76.22	585.03	418.29
0.97	210.16	62.35	25.66	94.76	86.84	672.32	490.57
0.98	246.38	73.35	30.17	110.67	102.05	797.78	595.80
0.99	309.36	92.56	38.05	138.32	128.55	1017.38	782.85

Table 7. (continued)

Prob.	(50,0)	(50,1,1)	(50,2,1)	(50,1,25)	(50,2,25)	(50,1,50)	(50,2,50)
0.01	4.63	5.48	8.52	3.35	33.19	4.72	23.26
0.02	3.70	4.31	6.64	2.56	26.31	3.71	18.86
0.03	3.16	3.65	5.57	2.13	22.37	3.13	16.32
0.04	2.79	3.18	4.83	1.83	19.62	2.73	14.55
0.05	2.51	2.83	4.28	1.60	17.52	2.43	13.19
0.06	2.28	2.55	3.83	1.42	15.83	2.19	12.10
0.07	2.09	2.32	3.46	1.27	14.42	1.99	11.18
0.08	1.92	2.12	3.14	1.15	13.22	1.82	10.40
0.09	1.78	1.95	2.87	1.04	12.17	1.68	9.72
0.10	1.66	1.80	2.63	0.95	11.25	1.55	9.11
0.15	1.20	1.25	1.77	0.62	7.83	1.09	6.87
0.20	0.90	0.90	1.23	0.43	5.58	0.79	5.38
0.25	0.68	0.67	0.86	0.30	3.98	0.60	4.30
0.30	0.53	0.50	0.61	0.23	2.80	0.46	3.50
0.35	0.41	0.38	0.44	0.19	1.93	0.37	2.89
0.40	0.33	0.31	0.34	0.18	1.30	0.32	2.44
0.45	0.27	0.27	0.30	0.19	0.88	0.30	2.11
0.50	0.25	0.27	0.31	0.23	0.63	0.31	1.90
0.55	0.24	0.30	0.37	0.29	0.55	0.35	1.79
0.60	0.26	0.35	0.49	0.38	0.65	0.41	1.80
0.65	0.30	0.45	0.66	0.49	0.95	0.51	1.91
0.70	0.38	0.59	0.91	0.64	1.46	0.65	2.17
0.75	0.49	0.78	1.25	0.83	2.25	0.83	2.59
0.80	0.66	1.05	1.71	1.08	3.42	1.09	3.24
0.85	0.90	1.43	2.37	1.43	5.18	1.45	4.23
0.90	1.29	2.02	3.37	1.95	7.97	2.00	5.85
0.91	1.40	2.18	3.65	2.09	8.74	2.14	6.31
0.92	1.52	2.36	3.95	2.24	9.62	2.31	6.83
0.93	1.67	2.57	4.31	2.42	10.65	2.51	7.43
0.94	1.84	2.82	4.72	2.63	11.86	2.73	8.14
0.95	2.04	3.12	5.22	2.88	13.32	3.01	9.01
0.96	2.29	3.49	5.84	3.19	15.15	3.34	10.10
0.97	2.63	3.97	6.66	3.59	17.56	3.79	11.54
0.98	3.11	4.67	7.82	4.16	21.06	4.42	13.63
0.99	3.97	5.88	9.86	5.16	27.24	5.53	17.34

Figure 5. Sample variance of the estimated LD values for the small sample size treatments.

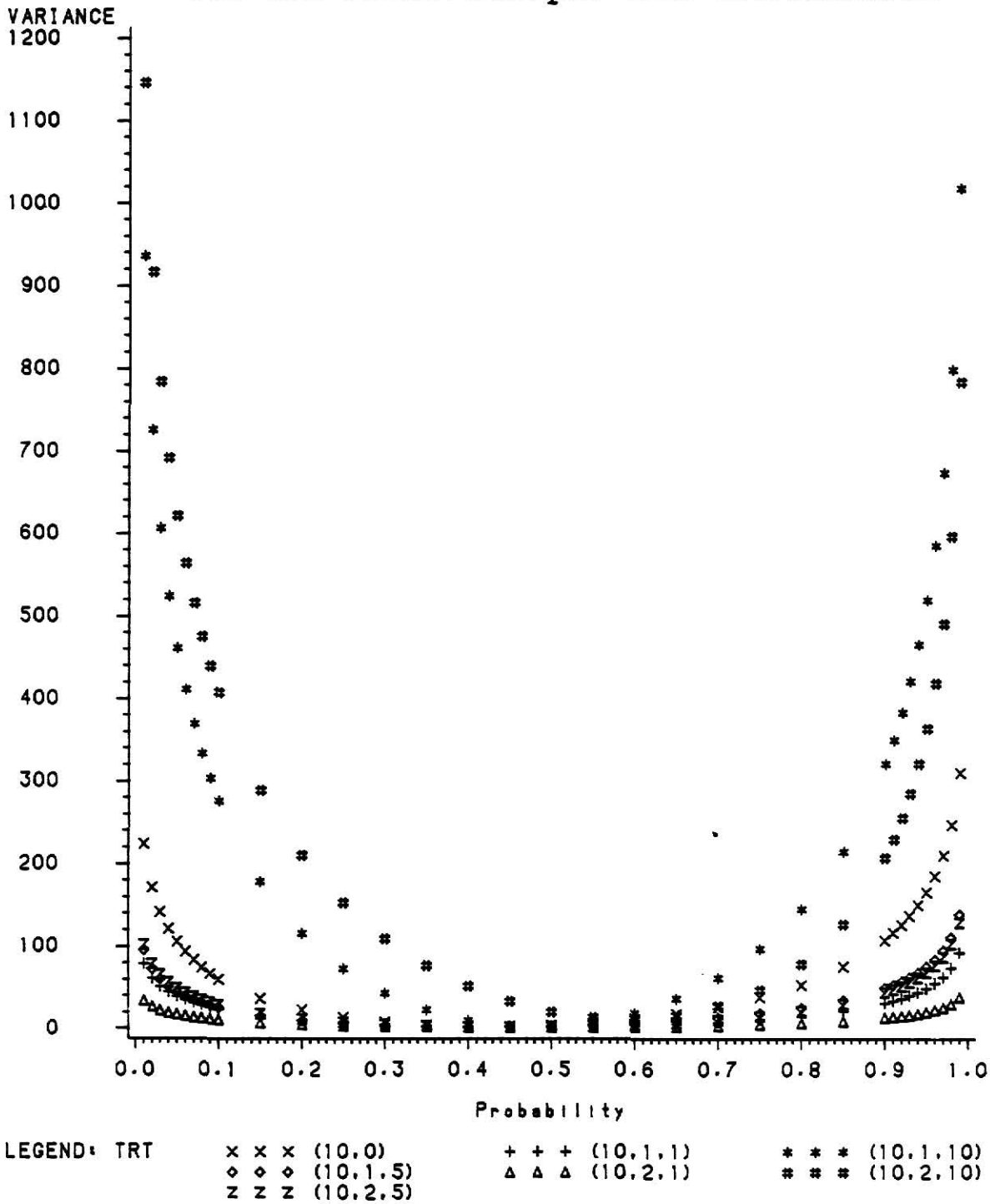
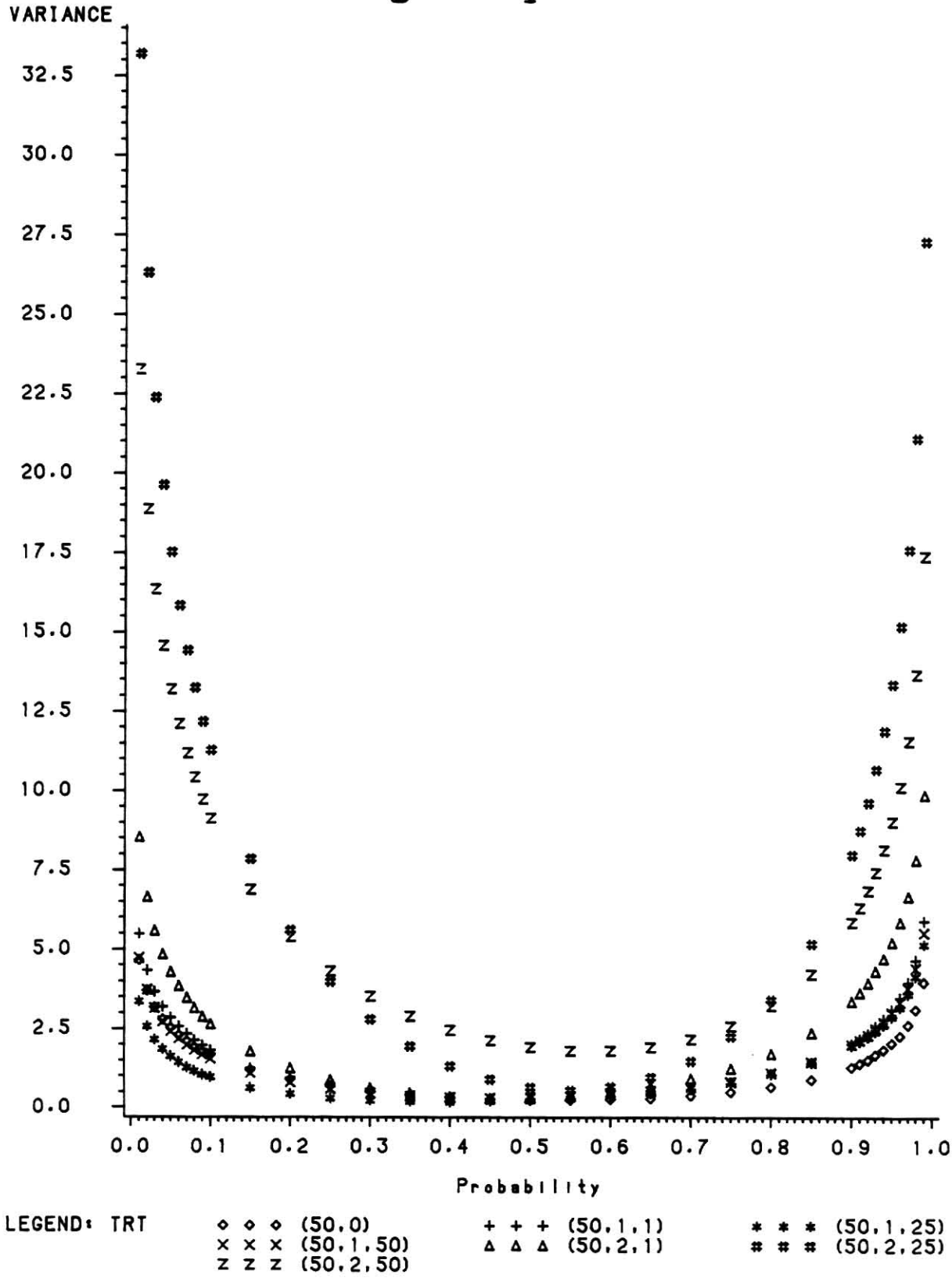


Figure 6. Sample variance of the estimated LD values for the large sample size treatments.



generally low around the LD50 value, which indicates that most of the predicted models are estimating the middle LD values with high precision. The precision varies dramatically, however, for the low and high percentile LD values.

In the large sample case, the nominal dose treatment has the smallest overall variance of the estimates. The estimates of the treatments with lower dosage errors have a precision that is fairly close to the nominal case. The higher dosage error treatments, however, are notably imprecise in the low and high percentiles. This indicates that extreme caution must be used when estimating a dose to which a high percentage of subjects would respond when the possibility of large dosage errors exist.

The estimates of LD parameters for the small sample treatments are also much less precise in the tails of the dose response curve. The sample variances indicate that while an individual application of the doses may not give precise estimates, at least the extreme responses which are possible by some subjects will be somewhat compensated for by other subjects in the group. When dosage errors affect the whole batch, however, so that all subjects given that treatment are uniformly affected, the responses may be so different from what is expected that individual experiments may not have reliable results. The precision of the estimates in the low and high percentiles for these treatments is very poor.

Table 7 and Figure 5 surprisingly show that the variance of the LD estimators for treatment (10,0), which has no dosage errors, is substantially higher than for each of the other small sample treatments which applied dosages either individually or to subgroups of the subjects. Figures 1 and 2 show that the distributions of the estimates of the model's mean and variance for treatment (10,0) are reasonable except for one extreme

outlier. This outlier causes the nominal dose treatment to look much worse than it is. When it is deleted and the sample variances of the LD estimators recalculated, the variances of treatment (10,0), as expected, become the lowest of all of the small sample treatments. This single outlier was judged to be a quirk of the random number sequence used in the simulation.

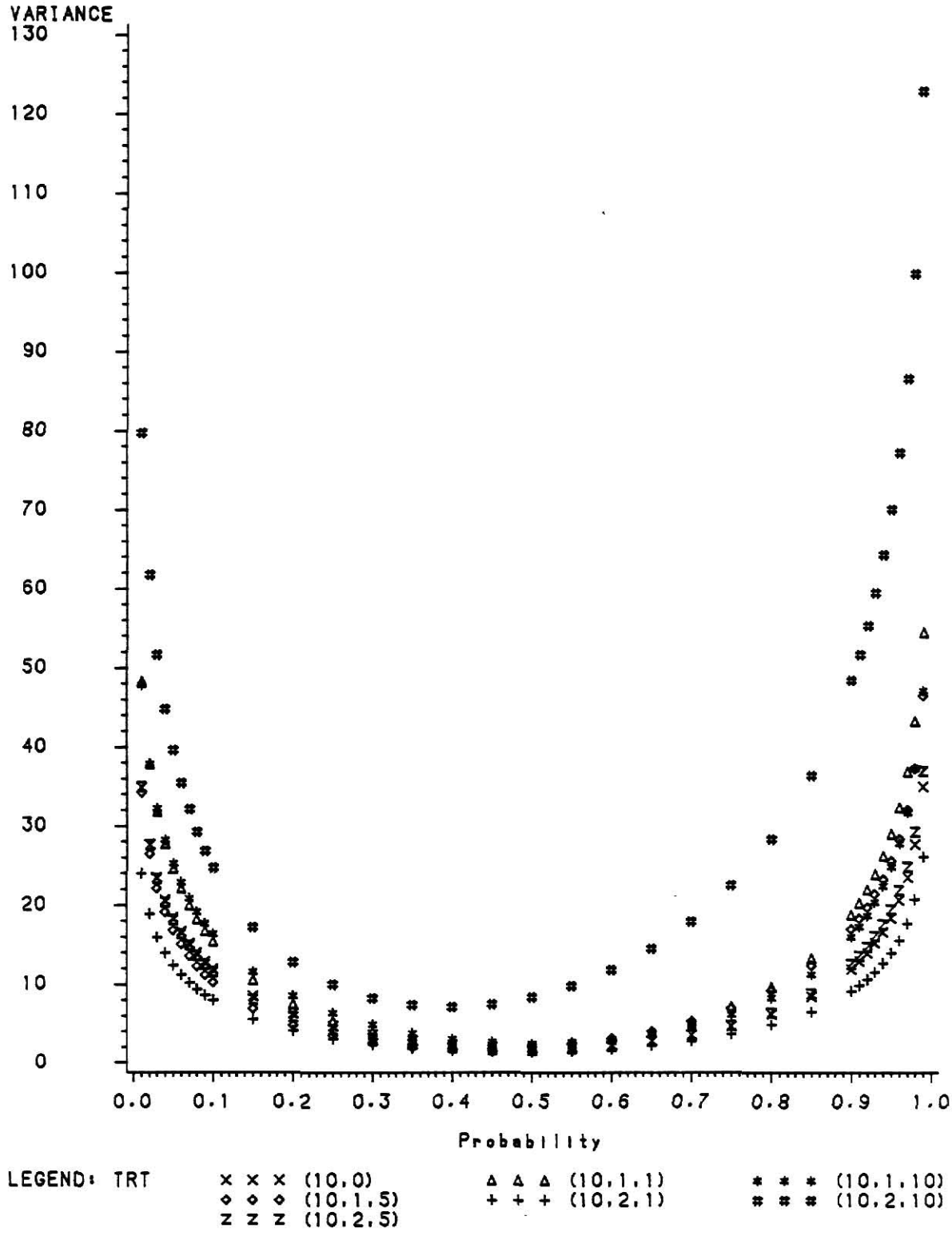
A second surprising feature of Table 7 and Figure 5 is how well treatment (10,2,1) does in comparison to treatment (10,1,1). With the higher dosage errors, treatment (10,2,1) would be expected to be less precise in estimating the LD values than (10,1,1), but it appears to be more precise. Table 6 indicates that treatment (10,2,1) provided estimates of the LD values which were closer, on the average, to the population values than any other treatment. That this is too good to be true is indicated by Table 8, which gives the sample variance of the LD estimators for a second run of 50 replications of treatment (10,2,1). Although the variances of the second run of (10,2,1) are still lower than (10,1,1) in the lower and upper percentiles, the variances are higher in the middle percentiles, around the LD50. Overall, the variances are within sampling errors at each percentile. This indicates a problem that may arise with the trustworthiness of the estimators for lower sample sizes. It appears that some of the estimates in treatment (10,1,1) may be poor enough to throw off the variance over all replications. At the same time, the first 50 replications of (10,2,1) resulted in a string of exceptional estimates of the model's parameters, a string that was not duplicated in a second run of this treatment.

To indicate the effect that the small sample cases may have on the maximum likelihood estimates, Figure 7 plots the variances of the LD estimators for the small sample treatments after the two worst replications

Table 8. Comparison of two runs of the (10,2,1) treatment: Sample variance of the LD values for each response probability.

Prob.	(10,1,1)	Run #1 (10,2,1)	Run #2 (10,2,1)
0.01	78.36	33.72	60.75
0.02	60.82	26.35	47.29
0.03	50.87	22.16	39.66
0.04	43.99	19.27	34.39
0.05	38.78	17.07	30.40
0.06	34.62	15.31	27.22
0.07	31.18	13.85	24.58
0.08	28.26	12.62	22.35
0.09	25.73	11.55	20.42
0.10	23.52	10.61	18.73
0.15	15.53	7.21	12.63
0.20	10.49	5.06	8.80
0.25	7.10	3.59	6.23
0.30	4.79	2.58	4.48
0.35	3.26	1.90	3.34
0.40	2.37	1.49	2.69
0.45	2.01	1.30	2.45
0.50	2.15	1.31	2.59
0.55	2.78	1.53	3.10
0.60	3.92	1.96	4.01
0.65	5.62	2.62	5.36
0.70	7.99	3.56	7.22
0.75	11.22	4.85	9.75
0.80	15.63	6.62	13.19
0.85	21.85	9.14	18.04
0.90	31.35	13.00	25.43
0.91	33.92	14.04	27.42
0.92	36.83	15.23	29.69
0.93	40.19	16.60	32.29
0.94	44.11	18.20	35.34
0.95	48.82	20.13	38.99
0.96	54.68	22.52	43.54
0.97	62.35	25.66	49.48
0.98	73.35	30.17	58.01
0.99	92.56	38.05	72.90

Figure 7. Sample variance of the estimated LD values for the small sample size treatments, with the 2 extreme values from each treatment deleted.



from each treatment were ignored. Although none of the other treatments have estimates quite as extreme as the one outlier of treatment (10,0), the small sample size treatments occasionallly produce very poor estimates. The result shown by Figure 7, compared to Figure 5, is a dramatic decrease in the variance of the estimators for all treatments over all percentiles. In fact, Figure 7 represents the range of variances expected for the small sample treatments based on the results of the large sample treatments in Figure 6. The differences in variances between most of the treatments are within sampling error, although the treatments which apply the doses collectively to the subjects still show considerably higher variances. In general, Figures 5 and 7 point out the problems that may occur in estimating the parameters of the probit model for experiments with small sample sizes.

The effects examined to this point indicate that the levels of factors considered in this study do not on the average have a significant effect on the bias of the estimates of the parameters of the probit model. A small sample size may cause the estimates of individual experiments to be wildly off, but on the balance the estimates will be reasonable. The precision of the estimates of the parameters, however, is significantly affected by the factors under consideration. Small sample sizes obviously hurt the precision of the estimates. The individual application of the doses to subjects evidently allows exceptional responses to even out within an experiment, while a collective application will sometimes throw off the results of an experiment making the overall estimates less precise. The presence of dosage errors does not singularly affect the estimates or their precision, but in combination with the other factors can hurt the precision of the estimates. Although the larger dosage error treatments slightly decrease the precision of the estimates, to this point no clear guidelines

are apparent as to how much dosage variation is tolerable. The worst case for an experimental situation demonstrated thus far is to apply dosages collectively to a small group of subjects when the dosage error is great enough so that successive dosages are within 2 standard deviations of error.

In order to further measure the effect of these related factors, the ability of the estimated 95% fiducial limits to contain the expected dosage level was monitored. For each percentile of the dose response curve within each replication, it was observed whether the expected dose fell within or without the estimated fiducial limits. The counts are presented in Table 9 in terms of the percentage of the replications for which the limits contained the expected dosage level.

A few general patterns are indicated by Table 9. The fiducial limits for the nominal dosage cases, treatments (10,0) and (50,0), are fairly good, especially for the larger sample size. The fiducial limits are not as accurate for the 12 dosage error treatments. The percentage included within the limits is particularly low for the dosage error cases when the subjects have been treated collectively as just one group. The limits for collectively applied dosages with a standard deviation of the errors of 2.0 (treatments (10,2,10) and (50,2,50)) were well below 95% limits at every percentile of the dose response curve. In general, when the 95% limits were deficient, they tended to fail in the middle percentiles, around the LD50 value. If the LD50 value is of primary interest in the study (and since, for a fixed number of subjects, dosages around the LD50 can usually be estimated more precisely than extreme LD values, the LD50 is often of primary importance), this is where it would typically be best for the fiducial limits not to fail.

Table 9. Percentage of the population LD percentiles within the 95% fiducial limits.

Prob.	(10,0)	(10,1,1)	(10,2,1)	(10,1,5)	(10,2,5)	(10,1,10)	(10,2,10)
0.01	98.0	98	100	94	94	94	84
0.02	98.0	98	100	94	94	94	84
0.03	98.0	98	100	94	94	94	82
0.04	98.0	98	100	94	94	94	82
0.05	98.0	96	100	94	94	94	82
0.06	98.0	96	100	94	94	94	82
0.07	98.0	96	100	94	94	94	82
0.08	98.0	96	100	94	94	94	82
0.09	98.0	96	100	94	94	94	82
0.10	97.5	96	100	94	94	94	82
0.15	96.0	98	98	96	94	94	86
0.20	96.0	98	98	96	94	92	86
0.25	96.0	98	98	98	94	94	84
0.30	94.5	98	98	98	96	96	84
0.35	95.0	96	100	98	96	94	80
0.40	94.0	94	100	98	96	94	78
0.45	93.5	94	100	98	96	92	76
0.50	93.0	90	96	96	96	94	72
0.55	94.5	92	94	96	98	94	76
0.60	93.5	90	90	92	100	94	82
0.65	92.5	92	92	92	96	94	84
0.70	94.5	94	94	92	96	92	86
0.75	95.5	94	94	92	96	92	92
0.80	96.0	94	94	92	96	92	92
0.85	95.5	92	94	92	96	92	90
0.90	95.0	94	92	92	96	92	90
0.91	95.0	94	92	92	96	92	90
0.92	95.5	94	92	92	96	92	90
0.93	95.5	94	92	92	96	92	90
0.94	95.5	94	92	92	96	92	90
0.95	95.5	94	92	92	96	92	90
0.96	95.0	94	94	92	96	94	88
0.97	95.0	94	94	92	96	94	88
0.98	95.0	94	94	92	96	94	88
0.99	96.0	94	94	92	96	94	88

Table 9. continued.

Prob.	(50,0)	(50,1,1)	(50,2,1)	(50,1,25)	(50,2,25)	(50,1,50)	(50,2,50)
0.01	95.5	98	88	100	86	100	90
0.02	96.0	98	88	100	86	98	90
0.03	96.0	98	88	100	86	98	90
0.04	96.0	98	88	100	86	98	90
0.05	96.0	96	88	100	86	98	90
0.06	96.0	96	88	100	86	98	88
0.07	96.0	96	88	100	86	98	88
0.08	96.0	96	88	100	86	98	90
0.09	96.0	96	90	100	86	98	90
0.10	96.0	96	90	100	88	96	90
0.15	95.5	96	92	100	88	96	90
0.20	95.5	98	94	100	88	96	90
0.25	95.5	98	94	100	88	96	88
0.30	96.0	98	94	98	88	96	86
0.35	96.0	98	94	98	88	94	88
0.40	94.5	92	94	98	94	92	84
0.45	95.0	92	96	98	96	92	80
0.50	97.5	92	96	100	98	92	84
0.55	97.5	96	94	100	98	92	84
0.60	97.5	96	92	98	98	92	86
0.65	97.5	92	94	96	94	94	88
0.70	98.5	94	94	96	92	96	90
0.75	98.0	94	94	96	92	98	88
0.80	98.5	96	92	96	90	98	88
0.85	98.0	96	90	96	90	98	90
0.90	98.0	96	92	96	90	98	90
0.91	98.0	96	92	96	90	98	90
0.92	98.0	96	92	96	90	98	90
0.93	98.0	96	92	96	90	98	90
0.94	98.0	96	92	96	90	98	90
0.95	98.0	96	90	96	90	98	88
0.96	97.5	96	90	96	90	100	88
0.97	97.0	96	90	96	90	100	90
0.98	97.0	96	90	96	90	100	90
0.99	97.0	96	90	98	88	100	92

A test of the effects of the experiment's factors on the ability of the 95% fiducial limits to contain the expected dosage level is somewhat problematic. The response of each replication at each percentile is dichotomous: the fiducial limits either contain the expected dose or they do not. Although analysis of variance procedures have often been used for dichotomous data when the probabilities of each response were close to .5, the data in the present case were presumptively constrained by probabilities of .05 and .95.

In a simulation study, Lunney (1970) has reported that a fixed effect analysis of variance procedure with 1-3 factors will give approximately correct results for dichotomous data. Considering probability levels of a success that ranged between .1 and .5, Lunney determined that the F-test is conservative for a small number of observations and that the Type I error and the power of the test are almost exactly controlled when the error degrees of freedom are large. For probabilities between .2 and .5, he suggested at least 20 error degrees of freedom, and recommended at least 40 for probabilities of .2 or less. The present data met these recommendations for sample size, although he did not consider probabilities as low as .05. Because of this, the cautions of Cochran and Cox (1957) about the approximateness of the significance levels of the F-tests must be heeded.

It was determined to conduct an analysis of variance to test the effects of the 3 experimental factors on the validity of the fiducial limits. An indicator variable was defined to represent the inclusion of the expected LD value within the fiducial limits at any given percentile of the probit model. This indicator variable is necessarily dichotomous since the expected LD is either included or not included in the fiducial limits. Thus, for each replication of the experiment and at each percentile of the

model, the indicator variable was given the value of 1 if those fiducial limits contained the expected LD value, and the value of 0 if the fiducial limits did not contain the expected LD. The analysis of variance was then conducted on this indicator variable over all 14 treatments.

Since the fiducial limits are not symmetric about a given LD value, and since the fiducial limits are also not symmetric for corresponding LD pairs (e.g., LD01 and LD99), it is difficult to combine information from several different LD percentiles. For the same reasons, it is also necessary to examine a fairly broad range of LD percentiles. Consequently, an analysis of variance was conducted for the fiducial limits at each .05 interval of the percentiles, as well as for LD01 and LD99. The analysis of variance tables, followed by a listing of pairwise t-tests on the treatment means, are given in Table 10.

The analysis on the fiducial limits of LD01 shows a significant interaction between sample size and the group size of the subjects. This interaction is due to the differences between the liberal fiducial limits of treatment (10,2,10), the moderate limits of treatment (10,2,5), and the conservative limits of treatment (10,2,1), which share a low sample size but used different application methods, combined with the lack of difference between the other treatment pairs that share either the same level of sample size or application but not both. Thus, for example, treatments (10,2,10), (50,2,25) and (50,2,50) each applied the dosages collectively while having different sample sizes, but the three treatments could not be significantly distinguished. The analysis of variance also shows an interaction between sample size and the level of dosage errors. The fiducial limits of the larger dosage error treatments are generally worse than the treatments with smaller dosage errors, but the exceptionally conservative limits of

Table 10. Analysis of variance tables on the inclusion within the fiducial limits of the population LD values for each percentile.

Analysis of variance for LD01

Source	DF	SS	F	PR > F
Sigma	2	0.93916667	10.19	0.0001
N	1	0.00810000	0.18	0.6752
Group size	2	0.16333333	1.77	0.1706
S*N	2	0.28488095	3.09	0.0460
S*G	2*	0.09000000	0.98	0.3771
N*G	2	0.36333333	3.94	0.0197
S*N*G	2*	0.14333333	1.55	0.2118
Error	986	45.45500000		
Total	999	47.50000000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	1.000	50	(50,1,50)
A	1.000	50	(50,1,25)
A	0.980	200	(10,0)
A B	0.980	50	(10,1,1)
A B	0.980	50	(50,1,1)
A B	0.955	200	(50,0)
A B C	0.940	50	(10,1,10)
A B C	0.940	50	(10,1,5)
A B C	0.940	50	(10,2,5)
B C D	0.900	50	(50,2,50)
C D	0.880	50	(50,2,1)
C D	0.860	50	(50,2,25)
D	0.840	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD05

Source	DF	SS	F	PR > F
Sigma	2	0.95433333	9.79	0.0001
N	1	0.00640000	0.13	0.7172
Group size	2	0.16333333	1.68	0.1878
S*N	2	0.20309524	2.08	0.1251
S*G	2*	0.16333333	1.68	0.1878
N*G	2	0.36333333	3.73	0.0244
S*N*G	2*	0.24333333	2.50	0.0829
Error	986	48.06000000		
Total	999	50.19100000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	1.000	50	(50,1,25)
A	0.980	200	(10,0)
A B	0.980	50	(50,1,50)
A B	0.960	200	(50,0)
A B C	0.960	50	(10,1,1)
A B C	0.960	50	(50,1,1)
A B C D	0.940	50	(10,1,10)
A B C D	0.940	50	(10,1,5)
A B C D	0.940	50	(10,2,5)
B C D E	0.900	50	(50,2,50)
C D E	0.880	50	(50,2,1)
D E	0.860	50	(50,2,25)
E	0.820	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD10

Source	DF	SS	F	PR > F
Sigma	2	0.70683333	7.19	0.0008
N	1	0.00090000	0.02	0.8924
Group size	2	0.26333333	2.68	0.0690
S*N	2	0.10988095	1.12	0.3272
S*G	2*	0.16333333	1.66	0.1902
N*G	2	0.25000000	2.54	0.0790
S*N*G	2*	0.24333333	2.48	0.0845
Error	986	48.43500000		
Total	999	50.19100000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	1.000	50	(50,1,25)
A	0.975	200	(10,0)
A B	0.960	200	(50,0)
A B C	0.960	50	(10,1,1)
A B C	0.960	50	(50,1,50)
A B C	0.960	50	(50,1,1)
A B C	0.940	50	(10,1,10)
A B C	0.940	50	(10,1,5)
A B C	0.940	50	(10,2,5)
B C	0.900	50	(50,2,50)
B C	0.900	50	(50,2,1)
C	0.880	50	(50,2,25)
D	0.820	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD15

Source	DF	SS	F	PR > F
Sigma	2	0.50016667	5.02	0.0068
N	1	0.00810000	0.16	0.6869
Group size	2	0.21000000	2.11	0.1220
S*N	2	0.06321429	0.63	0.5304
S*G	2*	0.08333333	0.84	0.4335
N*G	2	0.12333333	1.24	0.2904
S*N*G	2*	0.09000000	0.90	0.4055
Error	986	49.11500000		
Total	999	50.19100000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,25)
A B	0.980	50	(10,1,1)
A B	0.980	50	(10,2,1)
A B	0.960	200	(10,0)
A B C	0.960	50	(10,1,5)
A B C	0.960	50	(50,1,50)
A B C	0.960	50	(50,1,1)
A B C	0.955	200	(50,0)
A B C D	0.940	50	(10,1,10)
A B C D	0.940	50	(10,2,5)
A B C D	0.920	50	(50,2,1)
B C D	0.900	50	(50,2,50)
C D	0.880	50	(50,2,25)
D	0.860	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD20

Source	DF	SS	F	PR > F
Sigma	2	0.43516667	4.45	0.0119
N	1	0.00090000	0.02	0.8921
Group size	2	0.36333333	3.72	0.0246
S*N	2	0.08488095	0.87	0.4198
S*G	2*	0.07000000	0.72	0.4888
N*G	2	0.10333333	1.06	0.3477
S*N*G	2*	0.06333333	0.65	0.5232
Error	986	48.17500000		
Total	999	49.29600000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,25)
A B	0.980	50	(10,1,1)
A B	0.980	50	(10,2,1)
A B	0.980	50	(50,1,1)
A B	0.960	200	(10,0)
A B C	0.960	50	(10,1,5)
A B C	0.960	50	(50,1,50)
A B C	0.955	200	(50,0)
A B C D	0.940	50	(10,2,5)
A B C D	0.940	50	(50,2,1)
A B C D	0.920	50	(10,1,10)
B C D	0.900	50	(50,2,50)
C D	0.880	50	(50,2,25)
D	0.860	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD25

Source	DF	SS	F	PR > F
Sigma	2	0.66183333	6.82	0.0011
N	1	0.00250000	0.05	0.8205
Group size	2	0.44333333	4.57	0.0106
S*N	2	0.04202381	0.43	0.6487
S*G	2*	0.14333333	1.48	0.2289
N*G	2	0.08333333	0.86	0.4241
S*N*G	2*	0.06333333	0.65	0.5210
Error	986	47.85500000		
Total	999	49.29600000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,25)
A	0.980	50	(10,1,5)
A	0.980	50	(10,1,1)
A	0.980	50	(10,2,1)
A	0.980	50	(50,1,1)
A	0.960	200	(10,0)
A B	0.960	50	(50,1,50)
A	0.955	200	(50,0)
A B	0.940	50	(10,1,10)
A B	0.940	50	(10,2,5)
A B	0.940	50	(50,2,1)
B C	0.880	50	(50,2,50)
B C	0.880	50	(50,2,25)
C	0.840	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD30

Source	DF	SS	F	PR > F
Sigma	2	0.62983333	6.26	0.0020
N	1	0.02890000	0.57	0.4485
Group size	2	0.44333333	4.41	0.0124
S*N	2	0.10202381	1.01	0.3629
S*G	2*	0.20333333	2.02	0.1329
N*G	2	0.06333333	0.63	0.5329
S*N*G	2*	0.06333333	0.63	0.5329
Error	986	49.57500000		
Total	999	51.08400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.980	50	(10,1,5)
A	0.980	50	(10,1,1)
A	0.980	50	(10,2,1)
A	0.980	50	(50,1,25)
A	0.980	50	(50,1,1)
A	0.960	200	(50,0)
A B	0.960	50	(10,1,10)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,50)
A B	0.945	200	(10,0)
A B C	0.940	50	(50,2,1)
B C D	0.880	50	(50,2,25)
C D	0.860	50	(50,2,50)
D	0.840	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD35

Source	DF	SS	F	PR > F
Sigma	2	0.50733333	4.91	0.0076
N	1	0.00360000	0.07	0.7919
Group size	2	0.69333333	6.70	0.0013
S*N	2	0.03952381	0.38	0.6825
S*G	2*	0.25333333	2.45	0.0868
N*G	2	0.17333333	1.68	0.1876
S*N*G	2*	0.21333333	2.06	0.1276
Error	986	50.98000000		
Total	999	52.86400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	0.980	50	(10,1,5)
A	0.980	50	(50,1,25)
A	0.980	50	(50,1,1)
A	0.960	200	(50,0)
A B	0.960	50	(10,1,1)
A B	0.960	50	(10,2,5)
A B	0.950	200	(10,0)
A B	0.940	50	(10,1,10)
A B	0.940	50	(50,1,50)
A B	0.940	50	(50,2,1)
B C	0.880	50	(50,2,50)
B C	0.880	50	(50,2,25)
C	0.800	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD40

Source	DF	SS	F	PR > F
Sigma	2	0.24983333	2.07	0.1267
N	1	0.01210000	0.20	0.6544
Group size	2	1.04333333	8.65	0.0002
S*N	2	0.01916667	0.16	0.8532
S*G	2*	0.64333333	5.33	0.0050
N*G	2	0.09000000	0.75	0.4746
S*N*G	2*	0.10333333	0.86	0.4251
Error	986	59.49500000		
Total	999	61.64400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	0.980	50	(10,1,5)
A	0.980	50	(50,1,25)
A	0.960	50	(10,2,5)
A	0.945	200	(50,0)
A	0.940	200	(10,0)
A	0.940	50	(10,1,10)
A	0.940	50	(10,1,1)
A	0.940	50	(50,2,25)
A	0.940	50	(50,2,1)
A B	0.920	50	(50,1,50)
A B	0.920	50	(50,1,1)
B C	0.840	50	(50,2,50)
C	0.780	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD45

Source	DF	SS	F	PR > F
Sigma	2	0.27516667	2.25	0.1062
N	1	0.00010000	0.00	0.9678
Group size	2	1.71000000	13.97	0.0001
S*N	2	0.01059524	0.09	0.9171
S*G	2*	0.92333333	7.54	0.0006
N*G	2	0.06333333	0.52	0.5963
S*N*G	2*	0.02333333	0.19	0.8265
Error	986	60.35500000		
Total	999	63.37600000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,1)
A	0.980	50	(10,1,5)
A	0.980	50	(50,1,25)
A	0.960	50	(10,2,5)
A	0.960	50	(50,2,25)
A	0.960	50	(50,2,1)
A	0.950	200	(50,0)
A	0.940	50	(10,1,1)
A	0.935	200	(10,0)
A	0.920	50	(10,1,10)
A	0.920	50	(50,1,50)
A	0.920	50	(50,1,1)
B	0.800	50	(50,2,50)
B	0.760	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD50

Source	DF	SS	F	PR > F
Sigma	2	0.42983333	3.65	0.0263
N	1	0.20250000	3.44	0.0639
Group size	2	1.49333333	12.69	0.0001
S*N	2	0.07059524	0.60	0.5492
S*G	2*	1.05333333	8.95	0.0001
N*G	2	0.04000000	0.34	0.7120
S*N*G	2*	0.21333333	1.81	0.1638
Error	986	58.03500000		.
Total	999	61.64400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,25)
A B	0.980	50	(50,2,25)
A B	0.975	200	(50,0)
A B	0.960	50	(10,1,5)
A B	0.960	50	(10,2,5)
A B	0.960	50	(10,2,1)
A B	0.960	50	(50,2,1)
A B	0.940	50	(10,1,10)
A B	0.930	200	(10,0)
A B C	0.920	50	(50,1,50)
A B C	0.920	50	(50,1,1)
B C	0.900	50	(10,1,1)
C	0.840	50	(50,2,50)
D	0.720	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD55

Source	DF	SS	F	PR > F
Sigma	2	0.52233333	4.88	0.0078
N	1	0.11560000	2.16	0.1418
Group size	2	1.36333333	12.75	0.0001
S*N	2	0.02452381	0.23	0.7951
S*G	2*	0.56333333	5.27	0.0053
N*G	2	0.00333333	0.03	0.9693
S*N*G	2*	0.16333333	1.53	0.2177
Error	986	52.73000000		
Total	999	55.51900000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,25)
A	0.980	50	(10,2,5)
A	0.980	50	(50,2,25)
A	0.975	200	(50,0)
A	0.960	50	(10,1,5)
A	0.960	50	(50,1,1)
A	0.945	200	(10,0)
A	0.940	50	(10,1,10)
A	0.940	50	(10,2,1)
A	0.940	50	(50,2,1)
A B	0.920	50	(10,1,1)
A B	0.920	50	(50,1,50)
B C	0.840	50	(50,2,50)
C	0.760	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD60

Source	DF	SS	F	PR > F
Sigma	2	0.29766667	2.57	0.0773
N	1	0.12960000	2.24	0.1352
Group size	2	0.73000000	6.30	0.0019
S*N	2	0.02785714	0.24	0.7865
S*G	2*	0.42333333	3.65	0.0263
N*G	2	0.02333333	0.20	0.8178
S*N*G	2*	0.13000000	1.12	0.3264
Error	986	57.17000000		
Total	999	59.03100000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(10,2,5)
A B	0.980	50	(50,1,25)
A B	0.980	50	(50,2,25)
A	0.975	200	(50,0)
A B	0.960	50	(50,1,1)
A B	0.940	50	(10,1,10)
A B	0.935	200	(10,0)
A B C	0.920	50	(10,1,5)
A B C	0.920	50	(50,1,50)
A B C	0.920	50	(50,2,1)
B C D	0.900	50	(10,1,1)
B C D	0.900	50	(10,2,1)
C D	0.860	50	(50,2,50)
D	0.820	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD65

Source	DF	SS	F	PR > F
Sigma	2	0.23066667	1.88	0.1536
N	1	0.06760000	1.10	0.2945
Group size	2	0.20333333	1.65	0.1917
S*N	2	0.00142857	0.01	0.9884
S*G	2*	0.27000000	2.20	0.1117
N*G	2	0.00333333	0.03	0.9732
S*N*G	2*	0.07000000	0.57	0.5660
Error	986	60.59000000		
Total	999	61.64400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.975	200	(50,0)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,25)
A B	0.940	50	(10,1,10)
A B	0.940	50	(50,1,50)
A B	0.940	50	(50,2,25)
A B	0.940	50	(50,2,1)
B	0.925	200	(10,0)
B C	0.920	50	(10,1,5)
B C	0.920	50	(10,1,1)
B C	0.920	50	(10,2,1)
B C	0.920	50	(50,1,1)
B C	0.880	50	(50,2,50)
C	0.840	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD70

Source	DF	SS	F	PR > F
Sigma	2	0.35400000	3.36	0.0352
N	1	0.05760000	1.09	0.2960
Group size	2	0.12000000	1.14	0.3206
S*N	2	0.02666667	0.25	0.7765
S*G	2*	0.12000000	1.14	0.3206
N*G	2	0.05333333	0.51	0.6030
S*N*G	2*	0.05333333	0.51	0.6030
Error	986	51.95000000		
Total	999	52.86400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.985	200	(50,0)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,50)
A B	0.960	50	(50,1,25)
A B	0.945	200	(10,0)
A B C	0.940	50	(10,1,1)
A B C	0.940	50	(10,2,1)
A B C	0.940	50	(50,1,1)
A B C	0.940	50	(50,2,1)
A B C	0.920	50	(10,1,10)
A B C	0.920	50	(10,1,5)
A B C	0.920	50	(50,2,25)
B C	0.900	50	(50,2,50)
C	0.860	50	(10,2,10)

Table 10. (continued)

Analysis of variance for LD75

Source	DF	SS	F	PR > F
Sigma	2	0.29516667	2.99	0.0506
N	1	0.00810000	0.16	0.6853
Group size	2	0.03000000	0.30	0.7378
S*N	2	0.13821429	1.40	0.2467
S*G	2*	0.08333333	0.85	0.4298
N*G	2	0.00333333	0.03	0.9668
S*N*G	2*	0.07000000	0.71	0.4920
Error	986	48.61500000		
Total	999	49.29600000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.980	200	(50,0)
A	0.980	50	(50,1,50)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,25)
A	0.955	200	(10,0)
A B	0.940	50	(10,1,1)
A B	0.940	50	(10,2,1)
A B	0.940	50	(50,1,1)
A B	0.940	50	(50,2,1)
A B	0.920	50	(10,1,10)
A B	0.920	50	(10,1,5)
A B	0.920	50	(10,2,10)
A B	0.920	50	(50,2,25)
B	0.880	50	(50,2,50)

Table 10. (continued)

Analysis of variance for LD80

Source	DF	SS	F	PR > F
Sigma	2	0.47483333	4.93	0.0074
N	1	0.00250000	0.05	0.8199
Group size	2	0.02333333	0.24	0.7850
S*N	2	0.24892857	2.58	0.0761
S*G	2*	0.04333333	0.45	0.6380
N*G	2	0.01000000	0.10	0.9015
S*N*G	2*	0.03000000	0.31	0.7326
Error	986	47.51500000		
Total	999	48.39900000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.985	200	(50,0)
A B	0.980	50	(50,1,50)
A B	0.960	200	(10,0)
A B C	0.960	50	(10,2,5)
A B C	0.960	50	(50,1,25)
A B C	0.960	50	(50,1,1)
A B C	0.940	50	(10,1,1)
A B C	0.940	50	(10,2,1)
A B C	0.920	50	(10,1,10)
A B C	0.920	50	(10,1,5)
A B C	0.920	50	(10,2,10)
A B C	0.920	50	(50,2,1)
B C	0.900	50	(50,2,25)
C	0.880	50	(50,2,50)

Table 10. (continued)

Analysis of variance for LD85

Source	DF	SS	F	PR > F
Sigma	2	0.44416667	4.28	0.0140
N	1	0.01690000	0.33	0.5682
Group size	2	0.01000000	0.10	0.9081
S*N	2	0.24035714	2.32	0.0990
S*G	2*	0.04333333	0.42	0.6585
N*G	2	0.04333333	0.42	0.6585
S*N*G	2*	0.01000000	0.10	0.9081
Error	986	51.11500000		
Total	999	51.97500000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.980	200	(50,0)
A B	0.980	50	(50,1,50)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,25)
A B	0.960	50	(50,1,1)
A B	0.955	200	(10,0)
A B	0.940	50	(10,2,1)
A B	0.920	50	(10,1,10)
A B	0.920	50	(10,1,5)
A B	0.920	50	(10,1,1)
B	0.900	50	(10,2,10)
B	0.900	50	(50,2,50)
B	0.900	50	(50,2,25)
B	0.900	50	(50,2,1)

Table 10. (continued)

Analysis of variance for LD90

Source	DF	SS	F	PR > F
Sigma	2	0.40166667	3.87	0.0212
N	1	0.03240000	0.62	0.4298
Group size	2	0.01333333	0.13	0.8795
S*N	2	0.13500000	1.30	0.2730
S*G	2*	0.04000000	0.39	0.6804
N*G	2	0.04000000	0.39	0.6804
S*N*G	2*	0.04000000	0.39	0.6804
Error	986	51.20000000		
Total	999	51.97500000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.980	200	(50,0)
A B	0.980	50	(50,1,50)
A B	0.960	50	(10,2,5)
A B	0.960	50	(50,1,25)
A B	0.960	50	(50,1,1)
A B	0.950	200	(10,0)
A B	0.940	50	(10,1,1)
A B	0.920	50	(10,1,10)
A B	0.920	50	(10,1,5)
A B	0.920	50	(10,2,1)
A B	0.920	50	(50,2,1)
B	0.900	50	(10,2,10)
B	0.900	50	(50,2,50)
B	0.900	50	(50,2,25)

Table 10. (continued)

Analysis of variance for LD95

Source	DF	SS	F	PR > F
Sigma	2	0.56983333	5.41	0.0046
N	1	0.00810000	0.15	0.6949
Group size	2	0.02333333	0.22	0.8012
S*N	2	0.20488095	1.95	0.1433
S*G	2*	0.06333333	0.60	0.5481
N*G	2	0.02333333	0.22	0.8012
S*N*G	2*	0.02333333	0.22	0.8012
Error	986	51.89500000		
Total	999	52.86400000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	0.980	200	(50,0)
A B	0.980	50	(50,1,50)
A B C	0.960	50	(10,2,5)
A B C	0.960	50	(50,1,25)
A B C	0.960	50	(50,1,1)
A B C	0.955	200	(10,0)
A B C	0.940	50	(10,1,1)
A B C	0.920	50	(10,1,10)
A B C	0.920	50	(10,1,5)
A B C	0.920	50	(10,2,1)
B C	0.900	50	(10,2,10)
B C	0.900	50	(50,2,25)
B C	0.900	50	(50,2,1)
C	0.880	50	(50,2,50)

Table 10. (continued)

Analysis of variance for LD99

Source	DF	SS	F	PR > F
Sigma	2	0.49766667	4.99	0.0070
N	1	0.01960000	0.39	0.5310
Group size	2	0.00000000	0.00	1.0000
S*N	2	0.20738095	2.08	0.1257
S*G	2*	0.05333333	0.53	0.5862
N*G	2	0.12000000	1.20	0.3009
S*N*G	2*	0.09333333	0.94	0.3928
Error	986	49.20000000		
Total	999	50.19100000		

*Note: within levels of G, there are only two levels of Sigma.

Pairwise t-tests:
Means with the same letter are not significantly different.

Grouping	Mean	N	Treatment
A	1.000	50	(50,1,50)
A B	0.980	50	(50,1,25)
A	0.970	200	(50,0)
A B	0.960	200	(10,0)
A B C	0.960	50	(10,2,5)
A B C	0.960	50	(50,1,1)
A B C	0.940	50	(10,1,10)
A B C	0.940	50	(10,1,1)
A B C	0.940	50	(10,2,1)
A B C	0.920	50	(10,1,5)
A B C	0.920	50	(50,2,50)
B C	0.900	50	(50,2,1)
C	0.880	50	(10,2,10)
C	0.880	50	(50,2,25)

treatment (10,2,1) cause the dosage error effect to be masked in the small sample size case.

LD05 displays the same interaction between sample size and the method of application that LD01 had and for the same reason. The detrimental effect of greater dosage errors on the fiducial limits, which is shown strongly between the pair of treatments (10,1,10) (low dosage error) and (10,2,10) (high dosage error) and between (50,1,25) (low) and (50,2,25) (high), and shown moderately between treatments (50,1,50) (low) and (50,2,50) (high), is not matched at all by a similar dosage error effect between treatments (10,1,5) (low) and (10,2,5) (high). Thus the greater dosage errors harm the fiducial limits particularly when there is a small sample size and a collective application method.

The LD values for the percentiles from .10 to .35 do not demonstrate any significant interactions, but do show a highly significant dosage error effect, and a moderately significant group size effect. Treatments (10,2,10), (50,2,25) and (50,2,50), which each combine large dosage errors with a collective application, have the worst fiducial limits, typically by wide margins. In general the collective application of dosages causes a substantial detrimental effect. Large dosage errors then accentuate the problem for collective applications and may initiate the detrimental effects for individual application methods. Thus the 95% fiducial limits for the large dosage errors with a collective application generally contain only from 84-90% of the expected LD values.

The middle LD values LD40-LD60 each show a significant dosage error by group size interaction. For treatments with a collective application of the dosages, the fiducial limits of the treatments with the highest dosage errors are significantly worse than those from the treatments with lower

dosage errors. This dosage error effect, however, does not hold for the treatments with individually applied dosages. Treatment (10,2,1) almost singularly disturbs the pattern, since its fiducial limits contain an exceptionally high percentage of the expected LD values. Thus the combination of higher dosage errors with a collective application causes a detrimental effect on the accuracy of the fiducial limits.

The results for each of the probability levels from LD70 to LD99 are very similar. The fiducial limits around each of the dosage levels for these percentiles show only a significant dosage error effect. The treatments which have the greater variation around the dosages regularly have the lowest percentage of the expected dosages within their fiducial limits. In fact, from LD70 and up, none of the treatments with the greater dosage errors contain the expected LD value within their limits 95% of the time except for treatment (10,2,5). The treatments with the greater dosage errors and a collective application generally do not have even 90% limits.

In summary, the presence of dosage errors consistently affected the ability of the fiducial limits to contain the expected dosage level with the theoretical frequency. If the dosage errors were small enough that successive dosages were more than 2 standard deviations of the error apart, the fiducial limits often met the 95% inclusion rate. When less than 2 standard deviations of the error separated successive dosages, however, the fiducial limits are suspect, frequently not even containing the expected LD value for 90% of the replications. Often the method of application affected the seriousness of the problem. Thus, when the dosages were applied individually to each subject, the problem was sometimes not quite as serious. Usually, however, the presence of dosage errors was enough in itself to lower the confidence that can be placed in the fiducial limits.

A final point of interest is the effect of the experimental factors on the goodness of fit test performed on the estimated model. The goodness of fit test is designed to test the adequacy of the fit of the data to the probit model. If the test is made at the a100% significance level, it would be expected that (1-a)100% of the models would exhibit a good fit to the data. However, the presence of dosage errors or the method of application of the doses may affect the deviations of the data from the model and thus might affect this test of adequacy.

The SAS Probit procedure computes a 10% significance test. The percentage of the replications which met this test is summarized in Table 11 for each of the treatments. It is apparent that the nominal dosage treatments result in estimated models which fit the data with the prescribed frequency. The dosage error treatments do not do so well. The presence of the greater dosage errors usually causes the fit to be relatively poor. While the individually applied dose cases fit much better, when the doses are applied collectively to large groups of the subjects, and especially with large dosage errors, the estimated models fit the data quite poorly.

Table 11. Percentage of models which met the 10% goodness of fit test.

Treatment	# met/total	Percent	Treatment	# met/total	Percent
(10,0)	181/200	90.5	(50,0)	184/200	92
(10,1,1)	43/50	86	(50,1,1)	47/50	94
(10,2,1)	42/50	84	(50,2,1)	44/50	88
(10,1,5)	44/50	88	(50,1,25)	39/50	78
(10,2,5)	40/50	80	(50,2,25)	28/50	56
(10,1,10)	43/50	86	(50,1,50)	31/50	62
(10,2,10)	39/50	78	(50,2,50)	14/50	28

THE EFFECT OF RANDOM DOSAGES ON PROBIT ANALYSIS

by

BRUCE MORRILL

B. A., Kansas State University, 1977
M. A., Harding Graduate School, 1981

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1984

IV. CONCLUSION

The simulation of various levels of dosage variation, sample size, and the nature of the application of the doses indicates that these factors can affect the results of probit analysis. For the levels of factors considered in this study, the following general results are indicated:

1. The estimate of the mean of the probit model is unbiased over the levels of the factors considered in this study. However, the precision of this estimate is greatly hurt by a small sample size, and may also be hurt by large dosage errors when the dosages are applied collectively to the group of subjects.
2. The estimate of the variance of the probit model can be inflated by a small sample size and by large dosage errors. Again the precision of this estimate is fairly poor for small sample sizes and large dosage errors. As a result, the estimated LD values were too low for low response rates and too high for high response rates.
3. The estimated LD values appear to be unbiased. The variance of these estimates is small for the middle percentiles but is generally much larger for the extreme percentiles. For the small sample size case, the collective application of dosages wildly inflates the variance about the LD estimates, while an individual application of the dosages tends to balance out sporadic responses within an experiment, thus keeping the variance low. The levels of dosage errors considered in this study did not substantially affect the

precision of the estimates of the LD parameters in the small sample case unless the dosages were applied collectively. On the other hand, it is the level of dosage errors which affects the precision of the estimates in the large sample case, while the method of application has little effect. The variances for the smaller dosage errors were comparable to those for the no dosage error treatment but the variances were generally inflated for the large dosage error levels.

4. The estimated 95% fiducial limits are significantly affected by the presence of dosage variation, by the application of the same batch of dosages uniformly to each subject, or by some combination of these two factors. The fiducial limits in these cases are often too narrow and thus do not contain 95% of the population LD values.

Note that although the bias of the estimated mean is not affected on the average by these factors, the fiducial limits estimated about the LD50 value do not always merit 95% confidence. The limits for the cases with greater dosage variation and a collective application of the dosages may be too liberal.

The fiducial limits are quite good when there are no dosage errors. However, the fiducial limits are often suspect when the dosage variation is great enough so that successive dosage levels are within 2 standard deviations of each other. As the dose levels approach a separation of only 1 standard deviation of the dosage variation, the fiducial limits are generally poor. A smaller sample size study is more volatile, however, and especially so when the doses have been applied to the subjects as a group.

When the dosage variation is great and the doses are applied uniformly, it can play havoc with the fiducial limits. An application of the doses individually to each subject tends to somewhat balance out this problem.

The estimated variance of the probit model was inflated by the small sample size case. However, the larger estimated variance of the small sample size treatments made their respective fiducial limits broad enough to actually be 95% limits (averaged over the other factors). Meanwhile, the effect of the method of dose application and the effect of the dosage variation were masked in the estimation of the variance of the model, resulting in narrower fiducial limits for these factors. The fiducial limits then became too narrow and did not perform adequately for larger doasge variation and a collective application of the preparation.

Finney (1971) has indicated that the dosages should be selected so that a fairly broad response rate (e.g., 20-80%) is achieved. The results of this study indicate that if the dosage variation is not negligible, it would be best to ensure that each subject in the study receives an independent dose of the preparation rather than to treat a group of subjects with a uniform application of the preparation from the same batch. If the dosage variation is sufficiently controlled so that dosages are separated by more than 2 standard deviations of the dosage error, probit analysis appears to give satisfactory estimates.

REFERENCES

Cochran, William G. and Cox, Gertrude M. (1957). Experimental Designs. New York: John Wiley and Sons.

Finney, David J. (1978). Statistical Method in Biological Assay. London: Charles Griffin and Company LTD.

Finney, David J. (1971). Probit Analysis. Cambridge: University Press.

Lunney, Gerald H. (1970). 'Using Analysis of Variance with a dichotomous dependent variable: an empirical study,' Journal of Educational Measurement 7:263-69.

Milliken, George A. (1982). Nonlinear Statistical Models. Unpublished.

Patwary, Kamini M. and Haley, K. David C. (1967). 'Analysis of Quantal Response Assays with Dosage Errors,' Biometrics 23:747-60.

Snedecor, George W. and Cochran, William G. (1980). Statistical Methods. Ames, Iowa: Iowa State University Press.

ACKNOWLEDGEMENTS

I gratefully acknowledge my debt to Dr. Jim Higgins for suggesting and guiding this study. He is responsible for most of what is good in it.

I also appreciate the suggestions made by the committee members, Dr. George Milliken and Dr. John Boyer. My sincere appreciation and respect go to the faculty of the Department of Statistics for what they taught me in and out of the classroom about statistical work. I am grateful to the faculty for the assistantship which they provided and to Dr. Holly Fryer for a scholarship.

ABSTRACT

The dose response technique of probit analysis presumes that the dosages are known and fixed. This study examines the effects on the probit model of random variability in the dosages. Factors considered are the amount of variability in the dosages, the method of application of dosages (individual or collective), and the sample size. The estimates of the model parameters appear to be unbiased even for fairly large dosage errors. The precision of the estimates, however, is often suspect. Larger dosage errors and the administration of a faulty dose collectively to the subjects may seriously inflate the variance of the estimates. Fiducial limits generally do not contain the population parameters with the expected frequency when dosage errors are large and when the dosages are applied collectively.