OPTIMAL BIOASSAY DESIGN UNDER THE ARMITAGE-DOLL MULTI-STAGE MODEL

by

Christopher Jude Portier

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1362

October 1981

OPTIMAL BIOASSAY DESIGN UNDER THE ARMITAGE-DOLL MULTI-STAGE MODEL

by

Christopher Jude Portier

A Dissertation submitted to the faculty of The University Of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

October, 1981

Approved by:

Advisor

Reader

Reader

ABSTRACT

CHRISTOPHER PORTIER. Optimal Bioassay Design Under The Armitage-Doll Multi-Stage Model (under the direction of DAVID G. HOEL).

ABSTRACT: Laboratory based clinical trials are conducted to assess the toxicological effect of chronic exposure to a substance. Two distinct functions are generally addressed via this chronic animal bioassay; qualitative and quantitative risk assessment. The design of the chronic animal bioassay is an important issue which should be given careful consideration. Optimal designs aimed at quantitative risk assessment and satisfying standards for qualitative risk assessment are derived under the assumption that carcinogenesis can be modeled by the multi-stage theory of Armitage and Doll.

The distribution of the estimate of the virtually safe dose is derived using large sample theory. Designs are derived which minimize the asymptotic mean-squared-error of the estimate of the virtually safe dose. The linear, quadratic and linear-quadratic models are discussed in detail and general comments are made concerning the k-stage model. Asymptotically optimal designs for specific parameterizations of the multi-stage model are derived and compared to general designs such as that used by the National Cancer Institute.

The techniques of Monte Carlo simulation are used to locate designs which minimize the mean-squared-error of the estimate of the virtually safe dose for bioassays which use small numbers of animals. The distribution of the estimate of the virtually safe dose is discussed for small bioassays and compared to the distributions based on the large sample theory. The small bioassay optimal designs are compared to the asymp-

totically optimal designs as well as many general designs. Finally, the use of additional doses beyond the standard three is discussed with respect to improving the quantitative risk assessment for small bioassays.

The implications of the power of the test for positive tumor induction on optimal design selection are studied. Questions concerning the use of additional doses are explored as well as changes resulting from toxicity in the high-dose group. Recommendations are made concerning the design of the chronic animal bioassay.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express appreciation to my advisor, Dr. David G. Hoel, for his quidance and encouragement during the course of this research. I would also like to thank Dr. Michael Hogan and Dr. Robert Jennrich for their overall cooperation and advice. In addition, I also extend thanks to the other members of my committee, Dr. Dana Quade, Dr. Larry Kupper, Dr. Ronald Helms and Dr. Michel Ibrahim for their helpful comments.

The support for this research and most of my graduate study was through a P-appointment at the National Institute Of Environmental Health Sciences. I thank Dr. Hogan and Lois Thompson for their aid in obtaining this source of funding.

I wish to thank my wife, Beth, for the love and support she gave which enabled me to complete this research. I also appreciate the support given me by my family and friends during my graduate studies.

CHAPTER I

INTRODUCTION TO RISK ASSESSMENT

1.1 Risk Assessment

Quantitively assessing the risk of exposure to toxic chemicals in food and water is of increasing concern to man. In view of the exponential increase in the development and implementation of synthetic chemicals in recent years, a failure to accurately determine toxicological response associated with prolonged exposure to these chemicals could cause the human (or general) population to incur an unconscionable risk. Concern over the extended latency and irreversible nature of this toxicity make it imperative this risk be detected early in the development of a chemical. Of the various toxicological endpoints from such exposure, carcinogenesis has received the most attention. Two methods are commonly employed to ascertain cancer risk in human populations: epidemiological studies and animal based laboratory investigations.

Epidemiological studies are used to identify the occurrence of a disorder and its associated etiological factors under natural conditions in defined populations. Epidemiology uses the methods of population survey to discover the relationship between the occurrence of a disease and the presence of various biological, physical and social factors. Epidemiological studies are often inadequate for determining human response to environmental chemicals due to the lack of human cancer

data for most of the compounds being introduced. Because of the relatively long latency period, if we must wait for evidence of toxicity from epidemiological studies, we will have exposed the general population or some small occupational group to a possibly carcinogenic agent over an extended period with potentially adverse health effects.

In contrast to the natural environment of epidemiological studies, clinical studies assess the effect of a substance by setting up controlled environments and carefully monitoring exposure to the agent under study. Since it is not possible to perform a clinical study to determine cancer incidence in humans, clinical trials using animal models are conducted. The response for the animal model is extrapolated to estimate human carcinogenic potential (Campbell, 1980; Falk, 1980). At present, the main source of information on carcinogenic potential of environmental compounds is the chronic animal bioassay.

1.2 The Current Cancer Bioassay

One of the more common cancer bioassay designs currently in use in the United States is the design protocol used by the National Cancer Institute (NCI). The NCI cancer bioassay is usually performed on two species of animals: typically mice (C57BL/6 X C3H) and rats (F344). The species are grouped by sex into a two-by-two factorial design and the animals are followed over a period of 18 to 24 months.

The first step in any carcinogenesis bioassay is the estimation of the maximum tolerated dose (MTD). The MTD is defined (IRLG, 1979) as " ... the highest dose that can be administered to the test animals for their lifetime and that is estimated not to produce: a) clinical signs of toxicity or pathologic lesions other than those related to a

neoplastic response, but which may interfere with the neoplastic response; b) alteration of the normal longevity of the animals from toxic effects other than carcinogenesis; and c) more than a relatively small percent inhibition of normal weight gain (not to exceed 10%)". In actual practice, the MTD is estimated from a pre-chronic test over a wide range of doses with the restriction on weight gain being given the most considertaion.

Under the current NCI cancer bioassay, a series of doses is administered to 150 animals in each of the species-by-sex combinations. Using randomization by body weight, the animals are divided into three groups of 50 animals each; one group is a control group, the other two groups receive doses approximating the MTD and (generally) one-half the MTD for the compound being tested.

There are no serial sacrifices; the animals either die naturally or accidentally during the course of the study or are sacrificed at some predetermined termination date. The animals are necropsied at death and the target organs are usually subjected to both macroscopic and microscopic examination by a pathologist in order to evaluate toxic effects. In the typical cancer bioassay, data will include general survival information as well as type and number of tumors observed in specific tissues for each animal.

The information obtained from a cancer bioassay is generally used to fulfill two distinct functions: 1) to determine if there is an increased risk of cancer from exposure to the compound under study (i.e. qualitative risk assessment) ; and 2) to provide response information necessary for the assessment of potential human risks (i.e. quantitative risk assessment).

1.3 Qualitative Risk Assessment

There are two primary endpoints associated with the cancer bioassay; survival and tumor incidence. For survival analysis, the NCI recommends the use of Cox's life table techniques for pairwise comparisons and Tarone's extension of Cox's method for dose-response trends.

Lifetime tumor incidence data is often summarized like the hypothetical data in Table 1.1. In this instance, X_i represents the number of animals given dose d_i that have at least one tumor in the tissue under investigation. The total number of animals examined for tumors at dose d_i is n_i. Until recently, the analysis of this type of data was done using pairwise comparisons. In the NCI cancer bioassay, each of the two treated groups was compared to the control group for each species-by-sex combination with rejection of the hypothesis of no increased risk if the one-sided Fisher Exact Test p-value for the high dose group was less than .025. While this rule was not followed rigorously and other conditions were considered (e.g. the historical incidence of the tumor at interest for the animal model studied), this served as a "rule-of-thumb" in the interpretation of the test results.

TABLE 1.1. Sample Quantal-Response Toxicity Data

POSITIVE (tumor)	X_0	X_1	...	X_k
NEGATIVE	n_0-X_0	n_1-X_1		n_k-X_k
DOSE	d_0	d_1		d_k

More recently, in cases where several doses are used, the emphasis has been on the the use of trend tests. Trend tests are used to determine if increases in the dosage will result in an increased incidence. Armitage (1971) has proposed a test for linear trend in proportions which is commonly used in this situation. The test is based upon the proposition that the probability of an animal developing a tumor when administered d units of an agent is:

$$P(d)= \alpha + \beta d$$

where α and β are unknown parameters. The test is:

$$H_0: \beta=0 \quad \text{vs} \quad H_A: \beta>0$$

and is based upon the standard normal statistic:

$$Z= \{ \sum_i n_i (\hat{p}_i-\bar{p})(d_i-\bar{d}) \}\{ \bar{p}(1-\bar{p}) \sum_i n_i(d_i-\bar{d})^2 \}^{-.5} \qquad (1.3.1)$$

where:

$$\hat{p}_i = X_i / n_i \text{ for all } i \quad ;$$

$$\bar{p} = (\sum_i X_i) / (\sum_i n_i) \quad ;$$

and

$$\bar{d} = (\sum_i n_i d_i) / (\sum_i n_i) \quad .$$

One other procedure which is based only on incidence data was proposed by Chase and is discussed in Poon (1981). The Armitage linear trend test is the most widely used test for trends in incidence data.

1.4 Quantitative Risk Assessment

A great deal of discussion has appeared in the literature on what constitutes an acceptable risk and how to quantify it (Albert and Altshuler, 1969; Cornfield, et.al., 1978; Mantel and Bryan, 1961; Schneiderman, et.al., 1975; Van Ryzin and Rai, 1980; Wahrendorf, 1979). Generally, a quantity known as the virtually safe dose (VSD) is estimated. The VSD is that dose, s, which yields a very small, ε, increased incidence over the spontaneous background incidence. Commonly proposed values of ε are in the range of 10^{-8} to 10^{-4}.

Mantel and Bryan (1961) have shown that, in order to determine a very small added risk, it is necessary to use impractically large numbers of animals in the bioassay. However, the current bioassay is performed on a small number of animals using doses which yield a detectable risk over background of $>10^{-1}$ and we are forced to extrapolate from this "high" dose region to the "low" dose region to estimate the VSD. This process is commonly referred to as low-dose extrapolation. There are two methods generally employed to estimate the VSD from animal cancer bioassays: the safety factor approach and the mathematical dose-response model.

In the safety factor approach, the VSD is usually defined as some fraction of the highest dose for which no toxic effect was observed. This fraction or safety factor will range between 1/30 and 1/5000 depending on the toxicological endpoint and the quality of the data.

In the modelling approach, a dose-response model is fit to the bioassay data and the VSD is estimated by extrapolating in terms of this model to the exposure or risk level of interest. Commonly utilized models include the probit, logit, Weibull, multi-stage and

gamma multi-hit models. Relative comparisons will not be made among models and this research will be limited to estimation of the VSD using the multi-stage model format.

The multi-stage model is based on the assumption (Armitage and Doll, 1954) that the mechanism of carcinogenic response can be expressed as a series of k self-replicating stages, some of which are linearly related to dose and others which are spontaneous. The multi-stage model can be written as:

$$P(d) = 1 - e^{-(\sum_{i}^{k} \alpha_i d^i)} \qquad (1.4.1)$$

where k is a known positive integer and $\alpha_i \geq 0$ for all i, i=0,1,...k.

1.5 Notation

Consider a typical single animal bioassay in which N animals are divided into L+1 groups and given doses of a suspected carcinogen. In addition to the notation of Table 1.1, define:

γ_i: the proportion of animals allocated to the ith dose group ($\gamma_i N = n_i$) ;

P(d): the true but unknown model of dose-response behavior as it occurs in nature (the underlying dose-response model) ;

$\hat{P}(d)$: dose-response model estimated from the results of the bioassay ;

p_i: $P(d_i)$ ($q_i = 1-p_i$) ;

$\hat{p}_i$: $\hat{P}(d_i)$ ($\hat{q}_i = 1-\hat{p}_i$) ;

ε: incremental risk which is considered safe or acceptable (i.e. 10^{-8}, 10^{-6} and 10^{-4}) ;

s: (VSD) dose which satisfies $\varepsilon = P(s)-P(0)$;

$\hat{s}$: (estimated VSD) dose which satisfies $\varepsilon = \hat{P}(\hat{s})-\hat{P}(0)$.

This notation will be used in all subsequent discussion. Figure 1.1 portrays the notation defined above in terms of the dose-response structure.

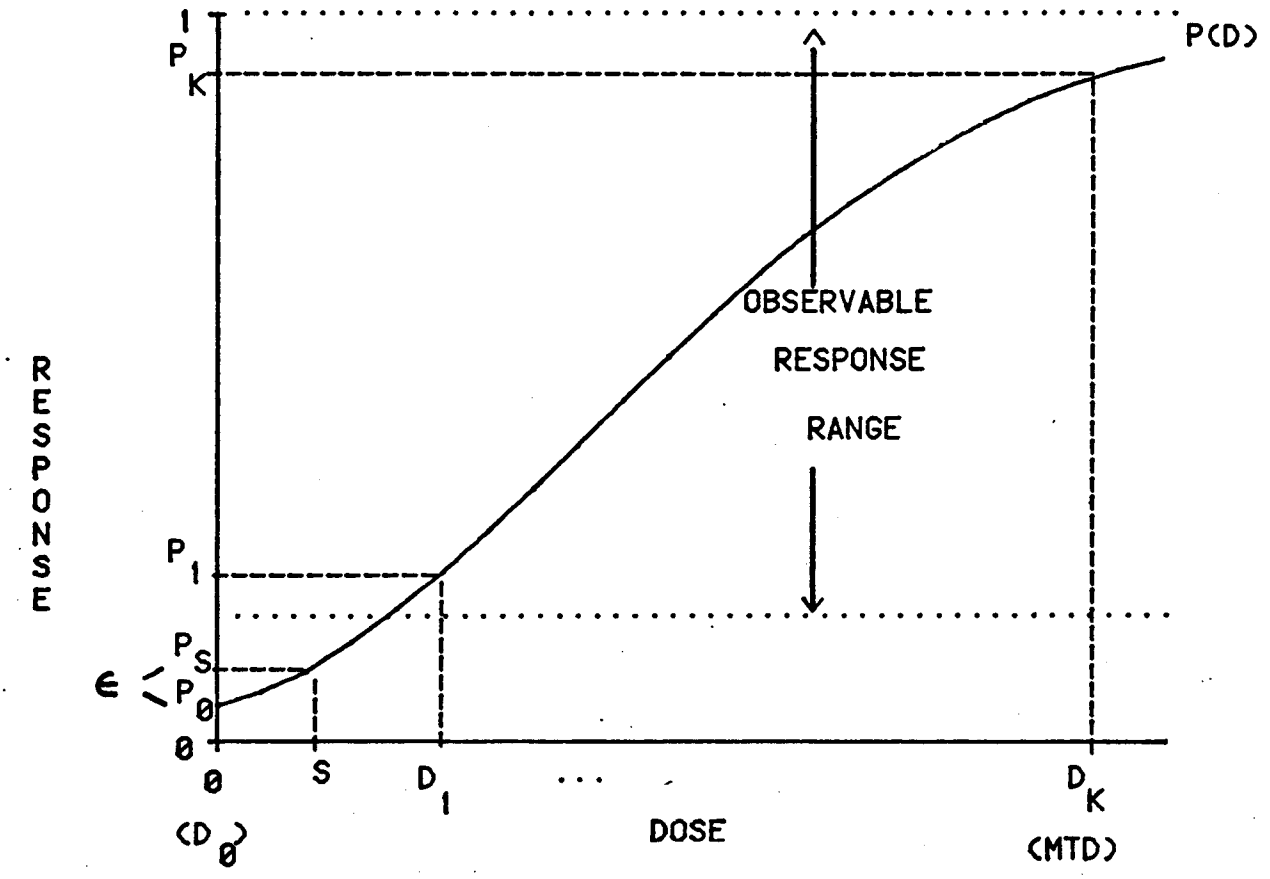
1.6 Statement Of The Problem And Outline

The purpose of this research is to consider what effect changes in the chronic animal bioassay will have on the power of the test for carcinogenicity and the estimation of the VSD. The main focus will be to change the dose and animal allocation structure to find that design which yields optimal estimation characteristics for the VSD while maintaining an acceptable level of power for detection.

In Chapter II the distribution of the estimate of the VSD for asymptotically large animal bioassays is discussed for the multi-stage model. Designs are derived which minimize the asymptotic mean-squared-error of the estimated VSD. The linear, quadratic and linear-quadratic models are discussed in detail and general comments are made concerning the k-stage model. Asymptotically optimal designs for specific parameterizations of the multi-stage model are found and compared to general designs such as the NCI protocol.

In Chapter III, the techniques of Monte Carlo simulation are used to find designs which minimize the mean-squared-error of the VSD for bioassays which use smaller numbers of animals. Again, the distribution of the VSD is discussed, this time for small bioassays, and compared

FIGURE 1.1: NOTATION



to the asymptotic distributions derived in Chapter II. Extensive work is done for the linear, quadratic and linear-quadratic models. These small sample optimal designs are compared to the asymptotically optimal designs as well as several other general designs. Finally, results are presented concerning the applicability of four-dose designs for improving the low-dose extrapolation.

In Chapter IV, the implications of the power of the test for positive tumor induction on optimal design selection is discussed. Questions concerning the use of additional doses (beyond the standard three) are explored as well as changes resulting from overestimation of the MTD.

Finally, in Chapter V, conclusions are drawn concerning design changes indicated by this study and some comments are made concerning suggestions for future research.

CHAPTER II

ASYMPTOTICALLY OPTIMAL DESIGNS

2.1 Optimal Designs

Let $\Omega = \{ \omega \}$ denote the set of all possible designs for a single animal bioassay, and define V to be a function from Ω onto the positive real numbers. A design, $\omega \varepsilon \Omega$, will be considered optimal if $V(\omega) \leq V(\omega')$ for all $\omega' \varepsilon \Omega$.

For the purposes of this chapter, V is defined to be the asymptotic mean-squared-error of the maximum likelihood estimate of the virtually safe dose, s. This will be referred to as the asymptotically optimal (AO) design.

Many authors have discussed the problem of choosing a design which minimizes the asymptotic variance of the VSD. Chernoff (1953) has shown that AO designs will have as many doses as there are unknown parameters in the model being used for extrapolation. Wong (1979) cites this result in her study of AO designs for the extreme-value model. By using information theory and reparameterizing the model, she was able to derive the asymptotic variance of the VSD. She found that, when the slope parameter is known, the optimal design will have all animals at that dose, d_1^0, satisfying $P(d_1^0) = .80$. When the slope parameter is unknown, the two optimal doses, d_1^0 and d_2^0, satisfy $P(d_1^0) = .10$ and $P(d_2^0) = .99$ with approximately 65% of the animals at dose d_1^0. Note

that, for this model and most subsequent models, the optimal design has no control dose since the model assumes the background response is known without error to be zero.

Krewski and Kovar (1982) used a second order approximation of the asymptotic variance of the VSD to derive designs which are optimal for the one-hit, probit, logit and extreme-value models with known slope parameter. For the one-hit and extreme-value models, they found the optimal dose, d_1^0, satisfies $P(d_1^0) = .80$ and, for the probit and logit models, $P(d_1^0) = .50$.

Using the same approach, Krewski, Kovar and Arnold (1981-1) discuss optimal designs for the probit, logit, Weibull, and gamma multi-hit models with known slope parameters. Table 2.1 is repeated from their article and gives the AO designs for different parameterizations of these models. In this paper, the authors also discussed designs in which the doses are restricted to be no greater than the MTD. They found that when the MTD is less than d_2^0, the unrestricted optimal upper dose, the optimal design then becomes d_2^0 = MTD and d_1^0 slightly smaller than its unrestricted counterpart, with 20% to 40% of the animals at the MTD.

Krewski, Kovar and Arnold (1981-2) have also discussed AO designs for the probit, logit, Weibull and gamma multi-hit models in the case of non-zero background response. After considering both additive and non-additive background incidence (Hoel, 1980), they conclude the AO three dose design will have 10% to 30% of the available animals at both the MTD and the control dose with the remaining animals being administered a relatively low middle dose when the background incidence was additive. For non-additive background, they observed the same animal allocations

TABLE 2.1: Two-Dose Optimal Experimental Designs

Model	Optimal Response Rates		Optimal Allocation (γ_1)			
	$P(d_1^0)$	$P(d_2^0)$	$\varepsilon = 10^{-2}$	10^{-4}	10^{-6}	10^{-8}
Probit	.058	.942	.84	.71	.67	.64
Logit	.083	.917	.76	.63	.59	.57
Weibull	.118	.971	.81	.72	.70	.68
Gamma Multi-Hit						
$\theta = 1$	.064	.947	.87	.80	.77	.76
2	.061	.944	.85	.76	.73	.71
5	.058	.943	.84	.73	.69	.67

reprinted from Krewski, Kovar and Arnold (1981-1)

but a slightly lower middle dose.

In the remainder of this chapter, the distribution of the VSD under the multi-stage model for asymptotically large (in terms of numbers of animals) bioassays is discussed. After deriving the distribution of the VSD, the mean-squared-error is derived and designs for minimizing this mean-squared-error are found.

2.2 Asymptotic Distribution Of The VSD

As previously discussed, the multi-stage model assumes:

$$P(d) = 1 - e^{-\sum_{i}^{k} \alpha_i d^i} \qquad (2.2.1)$$

for k a known positive integer and $\alpha_i \geq 0$ for all i, i=0,1,...k. Let $\hat{\alpha}_i$ be the maximum likelihood estimate of α_i, i=0,1,...k, and define the random vector Z by ;

$$Z' = (t_0, t_1, \ldots t_k) \qquad (2.2.2)$$

where $t_i = \sqrt{N}(\hat{\alpha}_i - \alpha_i)$.

Using the distribution of Z, it is possible to derive the distribution of the VSD. Define a positive definite matrix Σ by:

$$\Sigma^{-1} = [\ \sigma^{ij}\ ]$$

where :

$$\sigma^{ij} = (L+1) \sum_{l=0}^{L} \{(\gamma_l q_l)/p_l\}\ d_l^{\,i+j}\ . \qquad (2.2.4)$$

The elements of Σ will be denoted by σ_{ij}, i=0,1,...k, j=0,1,...k (For completeness, it is assumed $0^0=1$). Guess and Crump (1978) have shown that, if $\hat{\alpha}_i$ is allowed to range over the entire real line, then:

$$Z \rightarrow N_{k+1}(0,\Sigma) \ , \tag{2.2.4}$$

that is, Z is asymptotically distributed as a (k+1)-variate normal random variable with mean vector 0 and covariance matrix Σ.

The estimated VSD is defined by:

$$\varepsilon = \hat{P}(\hat{s}) - \hat{P}(0) \ .$$

Since $\hat{P}$, $\partial\hat{P}/\partial\hat{\alpha}_0$, $\partial\hat{P}/\partial\hat{\alpha}_1$, ... $\partial\hat{P}/\partial\hat{\alpha}_k$, $\partial\hat{P}/\partial\hat{s}$ are all continuous in any δ-neighborhood of $\alpha' = (\alpha_0, \alpha_1, \ldots \alpha_k)$ and since $\partial\hat{P}/\partial\hat{s}\Big|_{\hat{\alpha}=\alpha} \neq 0$, by the Implicit Function Theorem, there exists a continuous function g such that :

$$\hat{s} = g(\varepsilon,\hat{\alpha}) \ . \tag{2.2.5}$$

Furthermore, since $\hat{P}$ is well defined, both $\partial g/\partial\hat{\alpha}_i$ and $\partial^2 g/\partial\hat{\alpha}_i^2$ exist and are well defined. Let $B' = (b_0, b_1, \ldots b_k)$ be defined by:

$$b_i = \partial/\partial\hat{\alpha}_i \ g(\varepsilon,\hat{\alpha}) \Big|_{\hat{\alpha}=\alpha} \qquad \text{for } i=0,1,\ldots k. \tag{2.2.6}$$

Using Taylor's Theorem, $\hat{s}$ can be written as:

$$\hat{s} = g(\varepsilon,\alpha) + B'(\hat{\alpha}-\alpha) + O(1/N) \tag{2.2.7}$$

where O(1/N) is a function of $\hat{\alpha}$ of order 1/N. Then:

$$\sqrt{N}(\hat{s} - s) \rightarrow B'Z \tag{2.2.8}$$

which is denoted by:

$$M = B'Z \ . \tag{2.2.9}$$

The next step is to find the distribution of M when the maximum likelihood estimates, $\hat{\alpha}_i$, i=0,1,...k are constrained to be nonegative. The distributions can be classified into several cases depending on the value of α.

<u>Case 1:</u> $\alpha_i > 0$ for all i, $i=0,1,\ldots k$

Guess and Crump (1978) have shown that:

$$\lim_{N\to\infty} \Pr(\ \hat{\alpha}_i \leq 0\ ) = 0 \qquad (2.2.10)$$

so, by a corollary of that same paper;

$$Z \to N(\ 0\ ,\ \Sigma\ )\ . \qquad (2.2.11)$$

Using this fact and equation (25) (page 44) of Searle, it is clear that:

$$M \to N(\ 0\ ,\ B'\Sigma B\ )\ . \qquad (2.2.12)$$

<u>Case 2:</u> $\alpha_i=0$, $\alpha_j>0$ for all $j\neq i$, $j=0,1,\ldots k$

Define a matrix, $\Sigma_2 = [\ {}_2\sigma_{jl}\ ]$ by:

$${}_2\sigma_{jl} = \sigma_{jl} - \sigma_{ji}\sigma_{ii}^{-1}\sigma_{il}\ . \qquad (2.2.13)$$

In this case, it is true that:

$$\lim_{N\to\infty} \Pr(\ \hat{\alpha}_j \leq 0\ ) = 0 \quad \text{for } j\neq i,\quad j=0,1,2\ldots k\ .$$

Let ${}_u\hat{\alpha}_i$ denote the unconstrained maximum likelihood estimate of α_i and $\hat{\alpha}_i$ denote the constrained maximum likelihood estimate. If ${}_u\hat{\alpha}_i \geq 0$ then $\hat{\alpha}_i = {}_u\hat{\alpha}_i$ and the distribution of $\hat{\alpha}$ is unchanged from (2.2.11). However, when ${}_u\hat{\alpha}_i<0$, since the likelihood has normal contours, it will be maximized in the constrained space for $\hat{\alpha}_i=0$. The mapping used to go from the unconstrained space is linear and using equation (32) (page 47) in Searle, for this case:

$$Z \to N_{k+1}(\ 0\ ,\ \Sigma_2\ )\ .$$

So, in general:

$$Z \rightarrow \begin{cases} N_{k+1}(\ 0\ ,\ \Sigma\ ) & \text{for } \hat{\alpha}_i>0 \\ N_{k+1}(\ 0\ ,\ \Sigma_2) & \text{for } \hat{\alpha}_i=0\ . \end{cases} \tag{2.2.14}$$

(Note that Σ_2 is singular and the measure has non-zero mass (1/2) at $\hat{\alpha}_i=0$).

From (2.2.14) the asymptotic density of M can be written as:

$$f_M(m) = \int_{-\infty}^{0} f_M(m \mid z_i=t)\ f_{z_i}(t)\ dt \ + \int_{0}^{\infty} f_M(m \mid z_i=t)\ f_{z_i}(t)\ dt \tag{2.2.15}$$

where $f_M(m \mid z_i=t)$ is the density of M evaluated at m given $z_i=t$ and $f_{z_i}(t)$ is the marginal density of z_i (the ith element of Z) evaluated at t.

From (2.2.14), it follows that:

$$Pr(\ z_i=0\ ) = Pr(\ \sqrt{N}\ (\hat{\alpha}_i-\alpha_i)=0\ ) = Pr(\ \hat{\alpha}_i=0) = 1/2 \tag{2.2.16}$$

and;

$$Pr(\ z_i<0) = Pr(\ \sqrt{N}(\hat{\alpha}_i-\alpha_i)<0\ ) = Pr(\ \hat{\alpha}_i<0\ ) = 0 \tag{2.2.17}$$

Using (2.2.16) and (2.2.17), (2.2.15) can be rewritten as:

$$f_M(m) = (1/2)\ f_M(m \mid z_i=0\ ) + \int_{0}^{\infty} f_M(m \mid z_i=t\ )\ f_{z_i}(t)\ dt$$

From equation (27) (page 45) of Searle, asymptotically it is true that;

$$z_i \sim N(\ 0\ ,\ \sigma_{ii}\ )\ .$$

Using the same result as that cited for (2.2.12) and conditional normality arguments it follows that;

$$f_M(m \mid z_i=t) = \begin{cases} N(\ \mu t\ ,\ B'\Sigma_2 B\ ) & \text{for } t>0 \\ N(\ 0\ ,\ B'\Sigma_2 B\ ) & \text{for } t=0 \end{cases} \tag{2.2.18}$$

where $\mu = \sum^{k} b_1 \sigma_{1i} \sigma_{ii}^{-1}$. Simplifying this notation by using $\sigma_s^2 = B'\Sigma_2 B$, (2.2.15) can be rewritten as:

$$f_M(m) = (1/2)\ (2\pi\sigma_s^2)^{-\frac{1}{2}}\ \{e^{-\frac{1}{2}m^2/\sigma_s^2}\} +$$

$$\int_0^\infty (2\pi\sigma_s^2)^{-\frac{1}{2}}\ \{e^{-\frac{1}{2}(m-\mu t)^2/\sigma_s^2}\}\ (2\pi\sigma_{ii})^{-\frac{1}{2}}\ \{e^{-\frac{1}{2}t^2/\sigma_{ii}}\}\ dt \qquad (2.2.19)$$

Consider the second part of (2.2.19) given by:

$$\int_0^\infty (2\pi\sigma_s^2)^{-\frac{1}{2}}\ \{e^{-\frac{1}{2}(m-\mu t)^2/\sigma_s^2}\}\ (2\pi\sigma_{ii})^{-\frac{1}{2}}\ \{e^{-\frac{1}{2}t^2/\sigma_{ii}}\}\ dt \qquad (2.2.20)$$

This form can be simplified. The following algebra shows completion of the square for formula (2.2.20) and leads to a much simpler expression.

$$(2.2.20) = (2\pi\sigma_s^2\sigma_{ii})^{-\frac{1}{2}}\ \{\ e^{-\frac{1}{2}m^2/\sigma_s^2}\ \}\ \int_0^\infty (2\pi)^{-\frac{1}{2}}\ \{\ e^{-\frac{1}{2}(t^2/\sigma_{ii}\ -\ 2m\mu t/\sigma_s^2\ +\ \mu^2 t^2/\sigma_s^2\ )}\ \}\ dt$$

$$= (2\pi\sigma_s^2\sigma_{ii})^{-\frac{1}{2}}\ \{\ e^{-\frac{1}{2}m^2/\sigma_s^2}\ \}\ \int_0^\infty (2\pi)^{-\frac{1}{2}}\ \{\ e^{-\frac{1}{2}[(\sigma_{ii}\sigma_s^2)/(\sigma_s^2+\sigma_{ii}\mu^2)\ ]^{-1}\ [t^2-2t\{(m\mu\sigma_{ii})/(\sigma_s^2+\sigma_{ii}\mu^2)\}]}\ \}\ dt$$

$$= (2\pi\sigma_s^2\sigma_{ii})^{-\frac{1}{2}} \{ e^{-\frac{1}{2}m^2/\sigma_s^2} \} \int_0^\infty (2\pi)^{-\frac{1}{2}} \{ e^{-\frac{1}{2}[(\sigma_{ii}\sigma_s^2)/(\sigma_s^2+\sigma_{ii}\mu^2)]^{-1}}$$

$$[t^2 - 2t\{ (m\mu\sigma_{ii})/(\sigma_s^2+\sigma_{ii}\mu) \} + \{ (m\mu\sigma_{ii})/(\sigma_s^2+\sigma_{ii}\mu) \}^2] \}$$

$$\{ e^{-\frac{1}{2}[(m^2\mu^2\sigma_{ii})/(2\sigma_s^4+\sigma_s^2\sigma_{ii}\mu^2)]} \} \, dt$$

$$=(2\pi\sigma_s^2\sigma_{ii})^{-\frac{1}{2}} \{ e^{-\frac{1}{2}m^2/[\sigma_s^{-2} - (\mu\sigma_{ii})/(\sigma_s^4+\sigma_{ii}\mu^2)]} \}$$

$$\int_0^\infty (2\pi)^{-\frac{1}{2}} \{ e^{-\frac{1}{2}[(\sigma_{ii}\sigma_s^2)/(\sigma_s^2+\sigma_{ii}\mu^2)]^{-1}}$$

$$[t - (m\mu\sigma_{ii})/(\sigma_s+\sigma_{ii}\mu)]^2 \} \, dt$$

$$= (2\pi\sigma_s^2\sigma_{ii})^{-\frac{1}{2}} [(\sigma_{ii}\sigma_s^2)/(\sigma_s^2+\sigma_{ii}\mu^2)]^{\frac{1}{2}}$$

$$\{ e^{-\frac{1}{2}m^2/(\sigma_s^2+\sigma_{ii}\mu^2)} \} \, \Phi[(m\mu\sigma_{ii})/(\sigma_{ii}\sigma_s^2\{\sigma_s+\sigma_{ii}\mu^2\}^{\frac{1}{2}})] \quad (2.2.21)$$

where $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} dx$.

Expression (2.2.21) can be simplified by noting that:

$$\sigma_s^2+\sigma_{ii}\mu^2 = \sum_0^k \sum_0^k b_l b_j \, [\sigma_{lj}-\sigma_{li}\sigma_{ii}^{-1}\sigma_{ij}] + \sigma_{ii}[\sum_0^k b_j\sigma_{ij}\sigma_{ii}^{-1}]^2$$

$$= \sum_0^k \sum_0^k b_l b_j \sigma_{lj} - b_l b_j \sigma_{li}\sigma_{ii}^{-1}\sigma_{ij} + \sigma_{ii} \sum_0^k \sum_0^k b_l b_j \sigma_{lj}\sigma_{ii}^{-2}\sigma_{il}$$

$$= \sum_0^k \sum_0^k b_l b_j \sigma_{lj} = B'\Sigma B \qquad (2.2.22)$$

and;

$$\mu\sigma_{ii}^{\frac{1}{2}} = \sum_{0}^{k} b_l \sigma_{li} \sigma_{ii}^{-\frac{1}{2}} . \qquad (2.2.23)$$

Now (2.2.21) can be rewritten using (2.2.22) and (2.2.23) to yield:

$$(2\pi B'\Sigma B)^{-\frac{1}{2}} \{ e^{-\frac{1}{2}m^2/(B'\Sigma B)} \} \Phi [(m/\sigma_s\sigma) \Sigma b_l \sigma_{li}/\sigma_{ii}^{\frac{1}{2}}] \qquad (2.2.24)$$

where $\sigma = (B'\Sigma B)^{\frac{1}{2}}$.

Finally, the density of M using (2.2.19) and (2.2.24) is given by:

$$f_M(m) = \tfrac{1}{2}(2\pi\sigma_s^2)^{-\frac{1}{2}} \{ e^{-\frac{1}{2}m^2/\sigma_s^2} \} + \ (2.2.24) \ . \qquad (2.2.25)$$

<u>Case 3:</u> $\alpha_i=0$ for more than one i, i=0,1,...k

The aim of this research is to find modifications in the current single animal bioassay which will aid in the estimation of the VSD. Since the current animal bioassay incorporates only two dosed groups and a non-zero background is assumed, this case can be ignored as being of no interest to the problem under consideration. If this case is of interest to a particular reader, Guess and Crump (1978) discuss the asymptotic theory in more general terms.

2.3 Asymptotic Bias And Mean-Squared-Error Of The VSD

In trying to find an optimal design for estimating the VSD, the first approach would be to minimize the variance of the estimate of the VSD. However, in some cases, when dealing with constrained maximum likelihood, the estimate is biased. Thus minimizing the variance

would result in optimizing the design around the wrong point. Instead, it would be better to minimize around the true estimate of the VSD by minimizing the mean-squared-error given by:

$$MSE(\hat{s}) = E[(\hat{s} - s)^2] .$$

This situation applies to the asymptotic distribution of the VSD for some of the cases under study.

<u>Case 1:</u> $\alpha_i > 0$ for all i, i=0,1,...k

From (2.2.11) it is true that:

$$E(\hat{\alpha}_i - \alpha_i) = 0.$$

Using this it follows that:

$$\begin{aligned} E(\hat{s}) &= E[B'(\hat{\alpha}-\alpha)] + s \\ &= E[\sum_0^k b_i(\hat{\alpha}_i - \alpha_i)] + s \\ &= \sum_0^k b_i \; E(\hat{\alpha}_i - \alpha_i) + s \\ &= 0 + s = s \end{aligned}$$

and the estimate is unbiased. Similarly, from (2.2.11):

$$Var(\hat{\alpha}) = \Sigma$$

and using Searle:

$$(N)Var(\hat{s}) = B'\Sigma B .$$

Since, when $\hat{s}$ is unbiased, $Var(\hat{s}) = MSE(\hat{s})$, the asymptotic mean-squared-error of the VSD is given by:

$$(N)MSE(\hat{s}) = B'\Sigma B \quad . \tag{2.3.1}$$

<u>Case 2:</u> $\alpha_i=0$, $\alpha_j>0$ for all $j\neq i$, $j=0,1,\ldots k$

In this situation the distribution of Z is given by (2.2.14). It is known that:

$$E[N^{\frac{1}{2}}(\hat{\alpha}_j-\alpha_j)] = 0 \text{ for all } j\neq i \tag{2.3.2}$$

For the single parameter equal to zero, α_i, it is true that:

$$E[N^{\frac{1}{2}}(\hat{\alpha}_i-\alpha_i)] = \int_0^\infty z\ (2\pi\sigma_{ii})^{-\frac{1}{2}}\ e^{-\frac{1}{2}z^2}\ dz \tag{2.3.3}$$

After simple calculus, the following result is obtained;

$$E[N^{\frac{1}{2}}(\hat{\alpha}_i-\alpha_i)] = 2\sigma_{ii}^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}} \ . \tag{2.3.4}$$

From (2.3.2) and (2.3.4) it follows that:

$$E[N^{\frac{1}{2}}(\hat{s}-s)] = 2b_i\sigma_{ii}^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}} \ .$$

Thus, asymptotically, $\hat{s}$ is a biased estimator. Therefore, asymptotically, $VAR(\hat{s}) \neq MSE(\hat{s})$ and the optimal design will be found by minimizing the mean-squared-error of $\hat{s}$. For this case:

$$\begin{aligned}(N)MSE(\hat{s}) &= \tfrac{1}{2}B'\Sigma B + \tfrac{1}{2}B'\Sigma_2 B \\ &= B'\Sigma B - \tfrac{1}{2}\sum_0^k\sum_0^k b_l b_j \sigma_{li}\sigma_{ii}^{-1}\sigma_{ij} \ .\end{aligned} \tag{2.3.5}$$

For the remaining portion of this report, the asymptotically optimal design will be that design which minimizes formula (2.3.1) or (2.3.5) depending on the true parameter values, α_i.

Under an unconstrained multi-stage model, Hoel and Jennrich (1979)

have used the theory of Chebyshev systems to derive the asymptotic variance of the VSD. The formula which they optimize differs only by a constant from (2.3.1) and is given by:

$$V=\sum_{i=0}^{k} p_i(q_i\gamma_i)^{-1}\{(1-p_0)\prod_{\substack{j=0\\ j\neq i}}^{k}(-d_j)/(d_i-d_j) -$$

$$[1-P(s)]\prod_{\substack{j=0\\ j\neq i}}^{k}(s-d_j)/(d_i-d_j)\}^2 \qquad (2.3.6)$$

For Case 1 (all $\alpha_i>0$) this formula is equivalent to the mean-squared-error of $\hat{s}$ and can be used to find optimal designs. However, as seen in (2.3.5), it must be corrected when some $\alpha_i=0$ or it will overestimate the asymptotic mean-squared-error.

2.4 Two-Stage Models

Consider now the multi-stage model with only two stages;

$$P(d)=1-e^{-\alpha_0-\alpha_1 d-\alpha_2 d^2} \quad \alpha_0\geq 0, \alpha_1\geq 0, \alpha_2\geq 0 \qquad (2.4.1)$$

The three non-trivial forms of the two-stage model when non-zero spontaneous incidence is assumed are given by;

(a) the linear model; $P(d) = 1 - e^{-\alpha_0-\alpha_1 d}$ (2.4.2)

(b) the quadratic model; $P(d) = 1 - e^{-\alpha_0-\alpha_2 d^2}$ (2.4.3)

(c) the linear-quadratic model;

$$P(d) = 1 - e^{-\alpha_0-\alpha_1 d-\alpha_2 d^2} \qquad (2.4.4)$$

Assume the response which occurs in nature, P(d), is of one of these forms and a linear-quadratic model is used to estimate dose-response from a

sample obtained under P(d). In this case, it is possible to explicitly give the definition of g (2.2.5) as cases of the estimated values $\hat{\alpha}_i$:

$$\hat{s}=g(\varepsilon,\hat{\alpha})=\begin{cases}\{-\hat{\alpha}_1-[\hat{\alpha}_1^2-4\hat{\alpha}_2(\ln[e^{-\hat{\alpha}}0-\varepsilon]+\hat{\alpha}_0)]^{\frac{1}{2}}\}/(2\hat{\alpha}_2) & \text{for } \hat{\alpha}_0\geq 0,\ \hat{\alpha}_1\geq 0,\ \hat{\alpha}_2>0 \quad (2.4.4)\\ [-\ln(e^{-\hat{\alpha}}0-\varepsilon)-\hat{\alpha}_0]\ /\ \hat{\alpha}_1 & \text{for } \hat{\alpha}_0\geq 0,\ \hat{\alpha}_1>0,\ \hat{\alpha}_2=0 \quad (2.4.5)\\ \text{undefined elsewhere} & .\end{cases}$$

Calculating the partial derivatives of g and evaluating them for α, the following linear coefficients for (2.2.7) are derived.

<u>Case 1</u>: Underlying Linear Model

$$b_0= \varepsilon/[(\ e^{-\alpha}0 - \varepsilon)\ \alpha_1] \qquad (2.4.6)$$

$$b_1= C/\alpha_1^2 \qquad (2.4.7)$$

$$b_2= C/\alpha_1^3 \qquad (2.4.8)$$

where:

$$C= \ln(e^{-\alpha}0 - \varepsilon) + \alpha_0 \qquad (2.4.9)$$

<u>Case 2</u>: Underlying Quadratic Model

$$b_0= (-4\alpha_2 C)^{-\frac{1}{2}}\ \varepsilon/(e^{-\alpha}0 - \varepsilon) \qquad (2.4.10)$$

$$b_1= -1/(2\alpha_2) \qquad (2.4.11)$$

$$b_2= -(-C)^{\frac{1}{2}}/(2\alpha_2^{3/2}) \qquad (2.4.12)$$

where C is as defined by (2.2.9).

Case 3: Underlying Linear-Quadratic Model

$$b_0 = [\ \alpha_1^2 - 4\alpha_2 C\ ]^{-\frac{1}{2}}\ \varepsilon/(e^{-\alpha_0} - \varepsilon) \qquad (2.4.13)$$

$$b_1 = [\ \alpha_1\{\ \alpha_1^2 - 4\alpha_2 C\ \}^{-\frac{1}{2}} - 1]\ /\ (2\alpha_2) \qquad (2.4.14)$$

$$b_2 = [\ 2C\alpha_2 - \alpha_1^2 + \alpha_1(\alpha_1^2 - 4\alpha_2 C)^{\frac{1}{2}}]/[2\alpha_2^2(\alpha_1^2 - 4\alpha_2 C)^{\frac{1}{2}}] \qquad (2.4.15)$$

where C is as defined by (2.4.9) .

It is not practical to derive a closed form solution for the asymptotic mean-squared-error of the VSD. Instead, it is possible to numerically invert (2.2.3) and apply (2.4.6) through (2.4.15) to (2.3.1) and (2.3.5) to yield the mean-squared-error of $\hat{s}$. By performing a grid search over the practical designs, the optimal design for a given parameterization of P(d) can be found.

Using the theory of Chernoff, the number of dose groups in asymptotically optimal (AO) designs will equal the number of unknown parameters in $\hat{P}(d)$ (i.e. L=k). Mathematically, the MSE($\hat{s}$) is only affected by changing the doses (d_i) and the animal allocations (γ_i) when the total number of animals is fixed and (k+1) dose groups are used. Therefore, the AO designs presented will only include the dose and animal allocations.

2.5 Model Parameterizations

As previously mentioned, the AO designs can be located numerically for any given parameterization of the underlying (true-to-nature) dose-response model, P(d). By considering the historical dose-response observed from single bioassays, it is possible to derive a set of models which will span the range of two-stage models likely to occur in nature.

Gart, Chu, and Tarone (1979) have discussed the historical background incidence of cancer in laboratory animals. Table 2.2 is reprinted from their article and gives the percentage of animals with spontaneous tumors for the two species recommended for use in the NCI protocol and various target tissues. For the more commonly investigated tissues (e.g. liver, kidney, pituitary gland), spontaneous response can range from <1% to as high as 30%. Three background responses are considered in this study; .005, .05 and .20 .

In finding an optimal design, the highest dose is constrained to be less than or equal to the MTD. After studying the NCI Carcinogenesis Technical Report Series, it is clear that response at the MTD generally falls in the range of 30% to 50% with virtually all responses for positive carcinogens in the range 15% to 90%. For the purposes of determining optimal designs, four responses at the MTD are considered: 15%, 30%, 50% and 90%.

A linear model or a quadratic model can be uniquely defined by any two increasing dose-response points. By standardizing the dose scale to be MTD=1 and considering all practical combinations of background response and response at the MTD (3 backgrounds and four MTD responses as defined above), 11 parameterizations of the linear and quadratic models each are defined (note only 11 designs since background response of 20% and response at the MTD of 15% is not practical for a positive carcinogen). By choosing an appropriate third point a linear-quadratic model can be uniquely defined. Using a dose of ½MTD (d_1=.5) and response at this dose of ½ the MTD response [P(.5)=.5P(1)], 11 parameterizations of the linear-quadratic model are defined. Even though these points seem to follow a linear pattern, the parameterized linear-quadratic

TABLE 2.2: Percent Spontaneous Primary Tumors In Untreated Species Used At NCI For Carcinogen Bioassay

Organ Tissue	MICE (C57BL/6xC3H)F1		RATS F344	
	Male	Female	Male	Female
Brain	<1.0	-----	1.3	<1.0
Skin-Subcutaneous	1.0	<1.0	5.7	2.5
Mammary Gland	-----	<1.0	-----	18.8
Spleen	<1.0	<1.0	<1.0	<1.0
Lung-Trachea	9.2	3.5	2.4	<1.0
Heart	<1.0	-----	<1.0	<1.0
Liver	15.6	2.5	1.2	1.3
Pancreas	<1.0	<1.0	<1.0	-----
Stomach	1.1	<1.0	<1.0	<1.0
Intestines	<1.0	<1.0	<1.0	<1.0
Kidney	<1.0	<1.0	<1.0	<1.0
Urinary Bladder	-----	<1.0	<1.0	<1.0
Testis	<1.0	NA	76.2	NA
Ovary	NA	<1.0	NA	<1.0
Uterus	NA	1.9	NA	16.8
Pituitary Gland	<1.0	3.5	10.2	29.5
Adrenal Gland	<1.0	<1.0	8.7	4.0
Thyroid Gland	1.1	<1.0	5.1	5.6
Pancreatic Islets	<1.0	<1.0	3.2	1.3
Body Cavities	<1.0	<1.0	<1.0	<1.0
Leukemia-Lymphomas	1.6	6.8	6.5	5.4

reprinted from Gart, Chu, and Tarone (1979)

models range from near linear to near quadratic portraying the range of linear-quadratic models likely to be observed in nature. Table 2.3 gives the values of the parameters for these 33 models and the notation used to refer to them in later discussion. In Appendix A, these models are plotted as a function of dose and response.

As previously noted, much discussion has appeared in the literature concerning the definition of an acceptable risk. Formula (2.2.5) will be used to define the VSD and three acceptable risks are studied: 10^{-8}, 10^{-6}, and 10^{-4}.

2.6 Asymptotically Optimal Designs

When finding an optimal design for fitting a linear-quadratic model (2.4.4) to sample dose-response from an agent which follows one of the three two-stage model forms, (2.4.2) to (2.4.4), the asymptotic mean-squared-error of the estimated VSD, $\hat{s}$, is being minimized over five parameters; d_0, d_1, d_2, γ_0 and γ_2 (note that $\gamma_1=1-\gamma_0-\gamma_2$). By using a directed grid search (Gue and Thomas, 1968) it is possible to obtain the design which minimizes the mean-squared-error of $\hat{s}$. Table 2.4 gives the optimal designs for the 33 models of section 2.5 for an acceptable added risk of 10^{-6}. The doses are precise to within 10^{-3} of the optimal value and the allocations (γ_i) are within 10^{-4} of the optimal value. Appendix B gives the asymptotically optimal designs for acceptable risks of 10^{-8} and 10^{-4}.

From Table 2.4 and Appendix B it is seen that AO designs are robust with respect to the acceptable risk, ε. In fact, the choice of the AO doses seems to be completely independent of ε in this range. Also every AO design utilized a control dose ($d_0=0$), a practical necessity, and a

TABLE 2.3: Two-Stage Models Under Investigation

Back-ground	Model Type	MTD Response	Parameters α_0	α_1	α_2
.005	LINEAR	.15	.0050125	.15751	0.
		.30	.0050125	.35166	0.
		.50	.0050125	.68813	0.
		.90	.0050125	2.29760	0.
	QUADRAT	.15	.0050125	0.	.15751
		.30	.0050125	0.	.35166
		.50	.0050125	0.	.68813
		.90	.0050125	0.	2.29760
	LIN-QUA	.15	.0050125	.13429	.02322
		.30	.0050125	.27836	.07330
		.50	.0050125	.44254	.24559
		.90	.0050125	.07373	2.22380
.050	LINEAR	.15	.051293	.11123	0.
		.30	.051293	.30538	0.
		.50	.051293	.64185	0.
		.90	.051293	2.25130	0.
	QUADRAT	.15	.051293	0.	.11123
		.30	.051293	0.	.30538
		.50	.051293	0.	.64185
		.90	.051293	0.	2.25130
	LIN-QUA	.15	.051293	.01713	.09410
		.30	.051293	.13952	.16586
		.50	.051293	.30370	.33815
		.90	.051293	.31612	1.93520
.200	LINEAR	.30	.22314	.13353	0.
		.50	.22314	.47000	0.
		.90	.22314	2.07940	0.
	QUADRAT	.30	.22314	0.	.13353
		.50	.22314	0.	.47000
		.90	.22314	0.	2.07940
	LIN-QUA	.30	.22314	.12462	.00891
		.50	.22314	.23933	.23067
		.90	.22314	.22202	1.85740

TABLE 2.4: Asymptotically Optimal Designs For An Acceptable Added Risk Of 10^{-6}

Back-ground	Model Type	MTD Response	AO Designs d_0	d_1	d_2	γ_0	γ_1	γ_2
.005	LINEAR	.15	0.	.353	1.	.179	.648	.173
		.30	0.	.328	1.	.142	.685	.173
		.50	0.	.306	1.	.117	.710	.173
		.90	0.	.241	1.	.090	.742	.168
	QUADRAT	.15	0.	.270	1.	.293	.572	.135
		.30	0.	.225	1.	.267	.602	.131
		.50	0.	.189	1.	.247	.625	.128
		.90	0.	.125	1.	.218	.637	.145
	LIN-QUA	.15	0.	.315	1.	.198	.677	.125
		.30	0.	.284	1.	.161	.716	.123
		.50	0.	.263	1.	.142	.736	.122
		.90	0.	.134	1.	.195	.656	.149
.050	LINEAR	.15	0.	.448	1.	.322	.512	.166
		.30	0.	.401	1.	.285	.546	.169
		.50	0.	.367	1.	.255	.574	.171
		.90	0.	.280	1.	.218	.613	.169
	QUADRAT	.15	0.	.421	1.	.354	.507	.139
		.30	0.	.364	1.	.335	.520	.145
		.50	0.	.313	1.	.318	.532	.150
		.90	0.	.216	1.	.284	.535	.181
	LIN-QUA	.15	0.	.422	1.	.348	.514	.138
		.30	0.	.371	1.	.311	.553	.136
		.50	0.	.327	1.	.283	.580	.137
		.90	0.	.224	1.	.264	.564	.172
.200	LINEAR	.30	0.	.484	1.	.361	.481	.158
		.50	0.	.438	1.	.344	.493	.163
		.90	0.	.331	1.	.322	.511	.167
	QUADRAT	.30	0.	.465	1.	.369	.499	.132
		.50	0.	.412	1.	.352	.504	.144
		.90	0.	.296	1.	.321	.495	.184
	LIN-QUA	.30	0.	.466	1.	.364	.509	.127
		.50	0.	.414	1.	.348	.518	.134
		.90	0.	.297	1.	.320	.504	.176

group given the MTD. Although it is not shown, when the underlying dose-response model has response at the MTD of greater than .98, the AO designs do not utilize a group at the MTD. In every other case studied, the largest dose used in the AO design was the MTD.

It is possible to make some general remarks concerning AO designs after examining Table 2.4. As the response at the MTD increases, the optimal middle dose decreases and the proportion of animals to allocate to this group increases at the expense of the control group. This is misleading since the response for this optimal middle dose group is actually increasing as the MTD response increases, indicating there is no need to be as close to background if the agent being studied is not acutely toxic at small doses. Also note from Table 2.4, it is possible to bracket the optimal middle dose for an underlying linear-quadratic model to be less than that for a linear model and larger than that for a quadratic model using the same MTD response and background incidence. Thus, a distinct trend is shown in the optimal designs in going from linear to linear-quadratic to quadratic underlying models.

The NCI design protocol is not close to these optimal designs for any of the models being studied. The asymptotically optimal middle dose (d_1) ranges from 12.5% to 49% of the MTD and, in virtually every case, more than 50% of the animals are applied to this dose. This would indicate there are gains (possibly large gains) to be made in decreasing the mean-squared-error of the VSD by going to some other design protocol besides the NCI protocol.

The AO designs, as they appear in Table 2.4, are model dependent and the procedure for developing this kind of design can only be used when the response surface is known , without error, in advance. This,

as a practical approach to bioassay design, is unacceptable. However, a series of "fixed" designs can be compared to these optimal designs to find those designs which are nearly optimal for all of the models of underlying response. A "fixed" design is a design which is independent of the underlying model and based solely on the estimate of the MTD. These designs will be practical, since the only prior information on response for some agent to be studied deals with acute toxicity and the MTD. The NCI design is an example of a "fixed" design.

In Table 2.5 the asymptotic relative efficiencies (ARE) for several "fixed" designs (including the NCI design) as compared to the asymptotically optimal designs are given. The ARE of design ω' to the AO design, ω, is defined as:

$$ARE[\ \omega',\omega\ ] = V(\omega)/V(\omega')$$

for each underlying model and added acceptable risk, ε. For a design ω' to be considered as a replacement design for the NCI design, the $ARE[\ \omega',\omega\ ]$ must be close to 1 for every underlying model.

In Table 2.5, it is seen that the ARE for the NCI design ranges from .132 to .877. The NCI design seems to perform best when the underlying model is linear and does worst when the underlying model is quadratic. The NCI design also seems to perform poorly as compared to the AO design when the response at the MTD is high (>50%). In general, the trends observed for the NCI design are common in most designs. It is possible to reverse some of these trends, as shown by design II in Table 2.5. Design II reverses the loss in efficiency as the response at the MTD increases.

In choosing an optimal design which is based solely on the esti-

TABLE 2.5: Asymptotic Relative Efficiency Of Several "Fixed" Designs To The AO Designs

Back-ground	Model Type	MTD Response	Designs d_0,d_1,d_2 $\gamma_0,\gamma_1,\gamma_2$ NCI	II 0.,.2,1. .1,.7,.2	III 0.,.4,1. .3,.5,.2	IV 0.,.3,1. .2,.5,.3
.005	LINEAR	.15	.602	.654	.893	.901
		.30	.538	.781	.820	.893
		.50	.488	.862	.763	.870
		.90	.369	.962	.617	.820
	QUADRAT	.15	.457	.513	.787	.885
		.30	.341	.685	.625	.855
		.50	.262	.820	.493	.769
		.90	.132	.840	.240	.515
	LIN-QUA	.15	.521	.625	.833	.847
		.30	.448	.769	.752	.833
		.50	.385	.870	.667	.813
		.90	.146	.877	.264	.552
.050	LINEAR	.15	.847	.272	.962	.676
		.30	.775	.383	.990	.800
		.50	.699	.500	.971	.885
		.90	.513	.746	.813	.943
	QUADRAT	.15	.814	.215	.952	.167
		.30	.714	.288	.980	.735
		.50	.585	.385	.909	.847
		.90	.299	.680	.521	.877
	LIN-QUA	.15	.813	.219	.952	.606
		.30	.709	.322	.971	.752
		.50	.606	.439	.926	.847
		.90	.331	.714	.571	.901
.200	LINEAR	.30	.877	.208	.901	.568
		.50	.857	.282	.971	.683
		.90	.700	.471	.966	.877
	QUADRAT	.30	.833	.175	.885	.510
		.50	.805	.223	.962	.621
		.90	.541	.412	.840	.893
	LIN-QUA	.30	.826	.185	.885	.526
		.50	.803	.241	.959	.633
		.90	.552	.433	.855	.893

mated MTD, either of two strategies can be used. One strategy would be to find a "fixed" design which performs better than the NCI for all 33 underlying models under investigation. Of all the designs with ARE's larger than the NCI design, those for design III tend to be the largest (closest to 1). In design III, the MSE($\hat{s}$) ranges from 3% to 47% larger than that of the NCI design.

A second strategy would be to choose that design which yields the largest minimum ARE over all models. Design IV of Table 2.5 is that design. For the 33 models under investigation, the smallest ARE observed for design IV is .510 compared to .132 for the NCI design. However in comparison to the NCI design, this design can yield a mean-squared-error from 74% smaller than that for the NCI design to 63% larger.

In Table 2.5, it is seen that all three designs, II,III and IV, have a moderately low dose with a large proportion of the animals at that dose. Designs with this structure generally yielded the smallest mean-squared-errors. Even the AO designs tend to have this structure. The reason designs of this structure are optimal for extrapolation can only be hypothesized. When applying the linear-quadratic model, there is always a positive probability of fitting an incorrect model. The middle dose will determine whether the observed response is linear, quadratic or linear-quadratic. The response at control and the MTD group tend to be stable for relatively small numbers of animals. Therefore, response at the middle dose must be accurate enough to determine if the dose-response behavior is convex or concave in the low-dose region. This dose with most of the animals does that. In Chapter III, this reasoning will again follow.

The choice of an optimal design will depend on the needs of the researcher and the amount of prior information available. If from previous research on a given compound or similar compounds, the researcher has information available on response at the MTD or response at control (spontaneous incidence), he may wish to use this additional information to custom design a protocol for that compound. In this case, tables such as those in Appendix C are available. However, if general guidelines are to be employed and the only knowledge available is the estimated MTD, then design III or design IV should be utilized to make estimation of the VSD more accurate.

The results of this chapter are based upon the theory for asymptotically large bioassays. In the next chapter the discussion will center around optimal designs for small bioassays and a comparison of the small bioassay results to the asymptotic results of this chapter.

CHAPTER III

SMALL BIOASSAY OPTIMAL DESIGNS

3.1 Optimal Designs

All previous work on bioassay design has centered on the use of asymptotic theory to derive variance approximations for the estimate of the virtually safe dose. The distribution and properties of the VSD for small bioassays has never been investigated. A crucial assumption in most discussions utilizing asymptotic results concerns the distribution of the VSD for small bioassays. The intention of this chapter is to investigate the properties of the VSD for small bioassays and to develop designs which are optimal in this situation.

Let $\Omega_N=\{\omega_N\}$ denote the set of all possible designs for a single animal bioassay with a fixed total number of experimental units (animals) equal to N. A design, ω_N , is optimal for small bioassays if:

$$V(\omega_N) \leq V(\omega_N') \qquad (3.1.1)$$

for all $\omega_N' \varepsilon \Omega_N$ where V is the mean-squared-error of the constrained maximun likelihood estimate of the VSD. The design, ω_N, which satisfies (3.1.1) will be referred to as the small bioassay optimal (SO) design.

Guess, Crump, and Peto (1977) discussed the distribution of the model parameters of the multi-stage model for small bioassays and attempted to show what effect these distributions will have on the

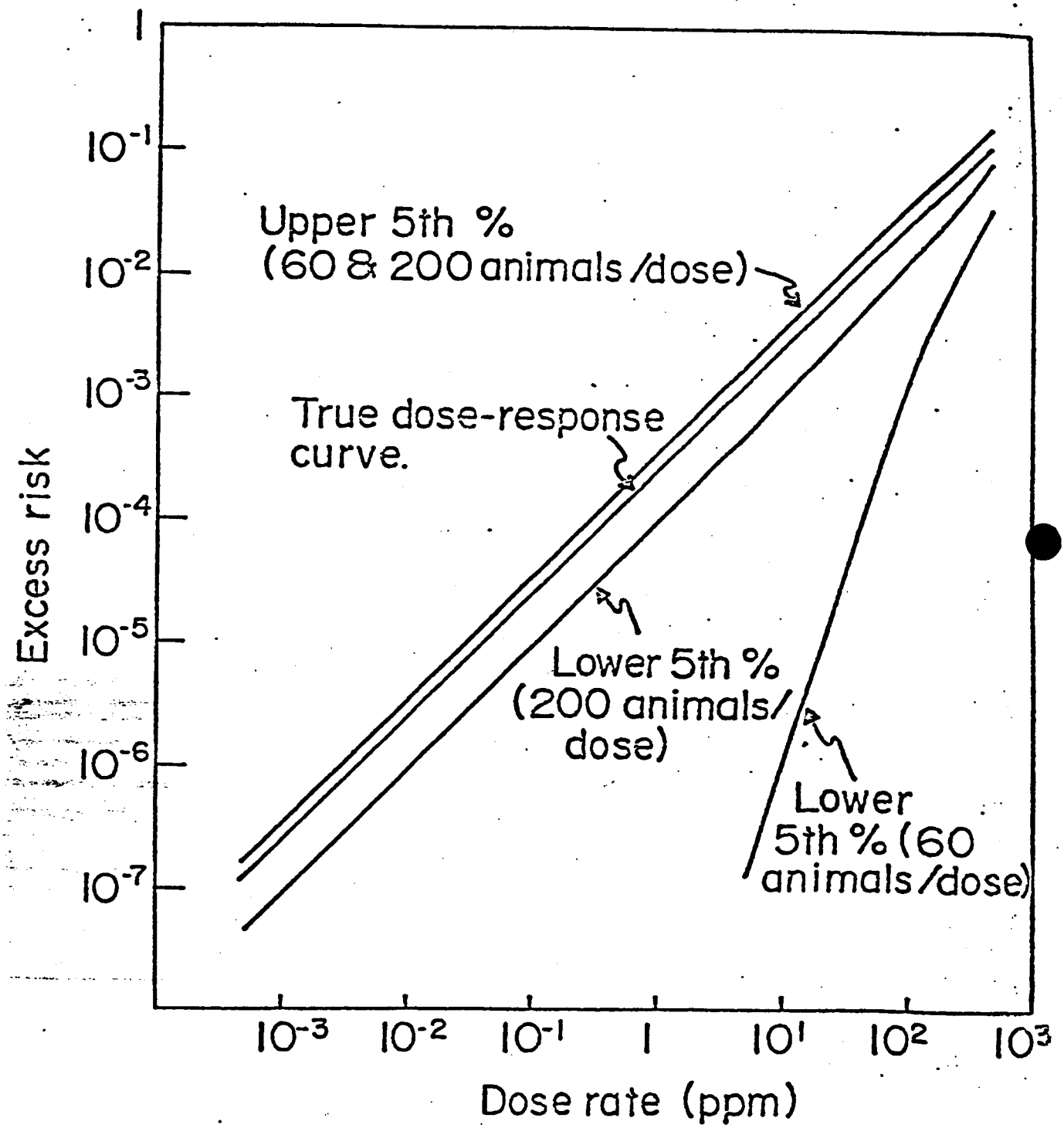
estimate of the VSD. Simulating experimental outcomes for several models from the literature, they produced graphs which depict the percentile range on the VSD. Figure 3.1 is a reprint from their article and shows the upper and lower 5% confidence bounds for the excess risk (ε) when, in nature, dose-response can be described by use of the multi-stage model:

$$P(d)= 1-e^{-.000267377d} \quad . \qquad (3.1.2)$$

The design used for this study was a four-dose design with doses of 0, 50, 250, and 500 (units are irrelevant to the discussion). The upper 5% confidence bound shown in Figure 3.1 is the same for both sets of animal allocations. However, for 200 animals per dose, the lower confidence bound is much closer to the true dose-response curve than the confidence bound for the experiment with 60 animals per dose. For a small excess risk ($<10^{-5}$), the lower confidence bound for excess risk uses a dose 3 to 4 orders of magnitude larger than that used for the true dose-response curve.

Although Figure 3.1 gives the confidence bounds for excess risk, these can be roughly inverted to yield a range on the VSD. In Figure 3.1, it is seen that, for small excess (or added) risks, the range on the VSD is quite wide. Changes in the design would, hopefully, decrease this range making the envelope curve smaller thus increasing the accuracy of the estimate of the VSD. The distribution of the VSD for small bioassays will be discussed in greater detail later in this chapter.

FIGURE 3.1


1
10-1
10-2
10-3
10-4
10-5
10-6
10-7
Excess risk
Upper 5th %
(60 & 200 animals /dose)
True dose-response
curve.
Lower 5th %
(200 animals/
dose)
Lower
5th % (60
animals /dose)
10-3
10-2
10-1
1
101
102
103
Dose rate (ppm)

3.2 Strategy and Methods

Many methods could be employed to calculate the mean-squared-error of the VSD for small bioassays. To avoid questions regarding the applicability of theoretically generated values of the mean-squared-error, Monte Carlo simulation is used to determine the mean-squared-error (Naylor, et.al., 1966).

To calculate the mean-squared-error of the estimate of the VSD, $\hat{s}$, it is necessary to know the true value of the VSD, s. This requires knowledge of the dose-response curve without error prior to conducting an experiment. To assure the general applicability of any one design as optimal, its optimality must be considered for a wide range of dose-response models. Due to constraints on time and expense, this investigation will focus on the use of the multi-stage model (2.2.1), in particular, the linear (2.4.1), quadratic (2.4.2) and linear-quadratic (2.4.3) forms of this model. No decisions concerning the correctness of using one model as compared to some other model are implied, but simply, the multi-stage model is chosen because of its wide acceptance and application.

Rather than base conclusions on models observed from experiments which have been conducted, the historic response in the cancer bioassay is again employed and the 33 models given in Table 2.3 are studied. It is again emphasized that these models portray a wide-range of dose-response behavior and are characteristic of the multi-stage models commonly observed from a bioassay.

Assume a bioassay experiment is designed, defining what doses (d_i) are to be used and how to allocate the animals (n_i) to the doses. If the dose-response relationship, P, is known in advance, it is possible

to derive the expected value of the binomial parameter at each dose: namely $p_i=P(d_i)$ for $i=0,1,\ldots k$. If a pseudo-random binomial deviate, x_i, with parameters p_i and n_i is generated for each dose, the x_i form a simulated experiment of what could occur in nature. By fitting a dose-response function, $\hat{P}(d)$, to this simulated response and inverting the function according to formula (2.2.5), the VSD, $\hat{s}$, can be estimated. Repeating this procedure R times, replicates of response for this experiment are created which form a random sample of the possible experiments. By fitting a model to each replicate and inverting the model, a random sample of the constrained maximum likelihood estimates of the VSD is created.

Altering the notation for the mean-squared-error of $\hat{s}$ to include the number of simulations, R, and the underlying dose-response model, P, for a fixed excess risk, ε, the mean-squared-error of $\hat{s}$ is given by:

$$V_N(\omega;R,\varepsilon,P) = (1/R) \sum_{r=1}^{k} (\hat{s}_r - s)^2 \qquad (3.2.1)$$

where $\hat{s}_r$ is the estimate of the VSD obtained from the rth replicate of the experiment. This is simply the standard formula for the mean-squared-error in a random sample of size R.

Since this is a random sample, it is known that for all r, $(\hat{s}_r - s_r)^2$ forms a set of mutually independent random variables, each with finite expectation. Therefore, by the strong law of large numbers, it follows that:

$$V_N(\omega;R,\varepsilon,P) \xrightarrow{a.s.} E\{(\hat{s}-s)^2\}. \qquad (3.2.2)$$

That is, the sample formula for the mean-squared-error of $\hat{s}$ (3.2.1) converges to the true mean-squared-error with probability 1 as R increases to infinity. Thus, by choosing R sufficiently large, (3.2.1) will yield a very good estimate of the mean-squared-error of $\hat{s}$. A preliminary study of the convergence properties of (3.2.1) was used to determine the value of R necessary for locating the SO designs. Using the NCI design and several of the models, the value of R was increased until (3.2.1) remained stable over repeated sets of R simulations for each model. The sample mean-squared-error seemed stable for 900 simulations, but to allow for any instability resulting from different designs and different models than those used in this preliminary study, each experiment was replicated 1000 times to estimate the mean-squared-error of $\hat{s}$.

It is impractical to exactly locate the optimal design for small bioassays. Not only is there still a small amount of variability in the estimate of the mean-squared-error of $\hat{s}$ for 1000 simulations, but the dose scale is continuous and, with 150 animals, all the possible allocations of animals to the doses would require impossible amounts of computer time. Instead, it is possible to find designs which are in a predetermined neighborhood of the truly optimal designs. Using a directed grid search (Bard, 1974), designs are found in which the doses are within .025 of the optimal doses and the animal allocations are correct to within 10 animals per dose. The criterion for optimality is now to find that design $\omega\varepsilon\Omega$ which satisfies:

$$V_N(\omega;R,\varepsilon,P) \leq V_N(\omega';R,\varepsilon,P) + \delta \qquad (3.2.3)$$

for all $\omega' \varepsilon \Omega$ and $\delta \geq 0$ very small.

By using the 33 models of Table 2.3, it is assumed some type of increasing dose-response relationship exists in nature. However, in simulating response for any one experiment using any one model, it is possible to create results in which no increasing dose-response relationship exists. In such cases, the determination of what constitutes a VSD is unclear. Therefore, these samples are censored to include only those experiments which have a linear trend significantly different from zero (defined as passing an Armitage linear trend test at the .05 significance level). Chapter IV will discuss the implications of this censoring and methods of incorporating this into the design problem.

3.3 Small Bioassay Distribution Of The VSD

When discussing the distribution of the VSD for small bioassays, Guess, et.al (1977) pointed out the broad range of the estimates yet failed to discuss why this range is so broad. Error in the estimation of the VSD can be the result of two related events. Performing a single bioassay for a small number of animals is, in essence, the realization of a random event. If response is obtained which is exceptionally different from the response occurring in nature, then the estimate of the VSD is likely to be wrong. Thus, error in the estimation of the VSD can be due to the randomness of the animal response.

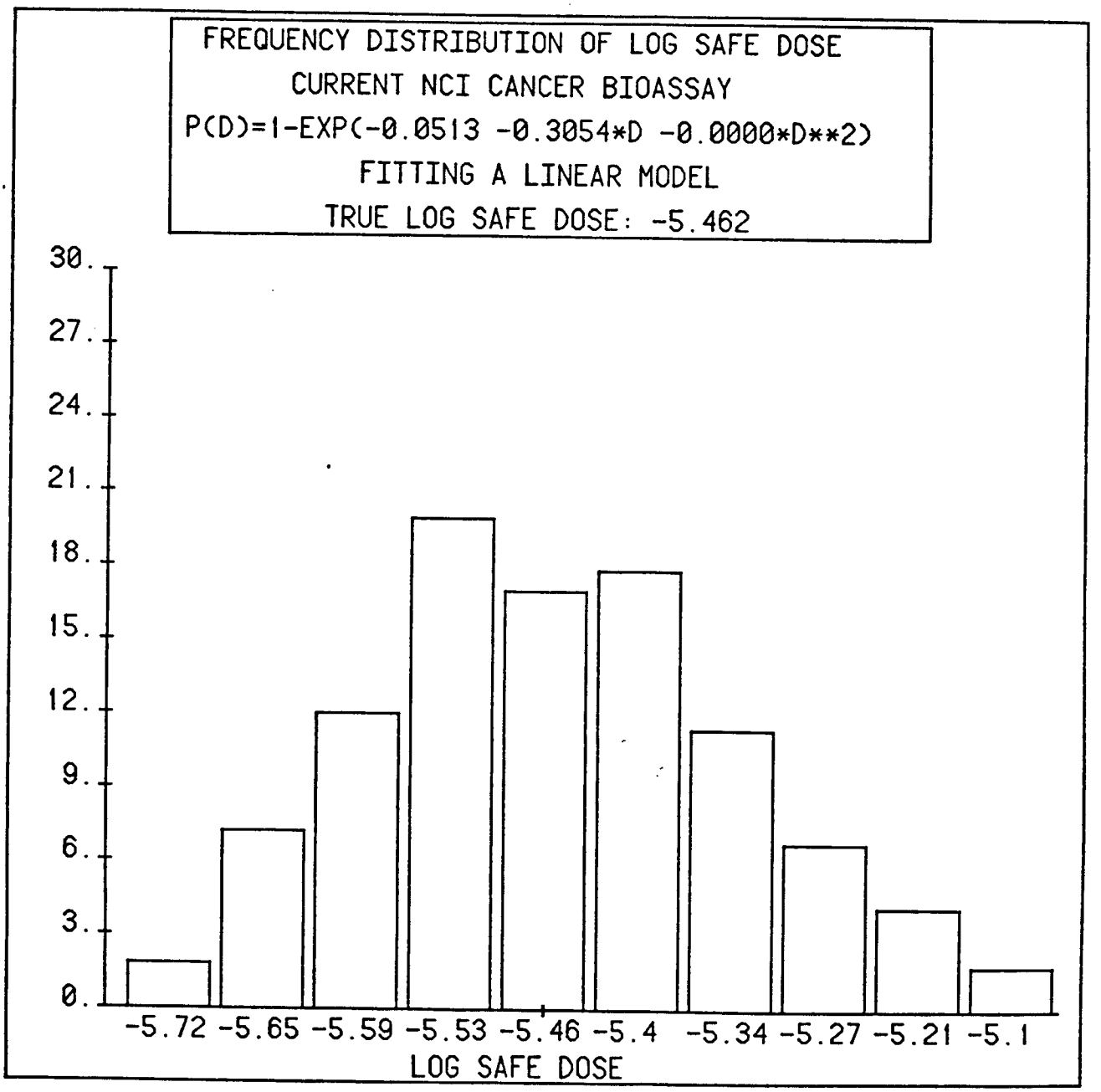
The second type of error is due to improper assumptions concerning the functional form of the dose-response relationship. In a single animal bioassay there are two models to be considered: the model which

the investigator fits to the response and the model which portrays the dose-response relationship as it exists in nature (the underlying model). If these two models are not the same, the estimate of the VSD could be drastically in error. In practice, assumptions are made concerning the form of the underlying model and some mathematical tool (such as maximum likelihood) is used to parameterize the model. In the investigation at hand, the underlying model is known without error and various models could be fit to simulated from this model response enabling determination of the effect of such model choice errors on the estimate of the VSD.

This discussion of the distribution of the VSD for small samples will begin by considering the simplest models. Assume that, in nature, the dose-response for some agent can be characterized by a linear model (2.4.1.). Under different assumptions, any number of functional forms could be fit to outcome from an experiment on this agent. Obviously, the correct model to fit to data of this type would be a linear model. Figure 3.2 is a frequency histogram of the estimates of the VSD (on the $\log_{10}$ scale) for this type of fitted model to underlying model relationship. The design used in this experiment was the NCI design and the underlying model is linear with .05 spontaneous response expected and an expected 30% response rate at the MTD. This distribution is unimodal and symmetric about the mean, possibly normal. The range on the estimates is small, from $10^{-5.8}$ to $10^{-5.1}$. Similar frequency histograms would result from the use of other designs and different parameterizations of the underlying linear model.

Suppose, in error, a quadratic model is fit to the outcomes from an

FIGURE 3.2:.


FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3054*D -0.0000*D**2)
FITTING A LINEAR MODEL
TRUE LOG SAFE DOSE: -5.462
30.
27.
24.
21.
18.
15.
12.
9.
6.
3.
0.
-5.72 -5.65 -5.59 -5.53 -5.46 -5.4 -5.34 -5.27 -5.21 -5.1
LOG SAFE DOSE

experiment conducted on an agent with this underlying linear model. The assumption used to fit this quadratic model is quite different from what is truly occurring in nature. In the linear model, the slope of the distribution at a dose of zero (control) is greater than zero. For the quadratic model, the slope of the distribution at control is zero. Thus, when ε is very small, the difference in the estimate of the VSD for the two models will be drastic. These points become clearer when we examine Figure 3.3. Although the models used in Figure 3.3 are obviously fabricated, they illustrate the point concerning the slope of the model at control. This figure shows that the linear and linear-quadratic models yield much smaller estimates of the VSD than the quadratic model.

In terms of fitting a quadratic model to response which is linear in nature, it is expected that the estimates of the VSD will be larger than the expected value. Figure 3.4 illustrates what the distribution of the estimates will look like in this case. Here, the distribution is again close to normal and covers a small range of possible estimates. However, in this case the mean of the estimates is approximately $10^{-2.7}$, over two orders of magnitude larger than the true VSD, $10^{-5.462}$. Thus, by choosing to fit a quadratic model, the risk from exposure to this agent has been underestimated.

Now, suppose the underlying model is quadratic. By fitting a quadratic model to realizations from an experiment with this underlying response, a correct assumption would be made. The estimates of the VSD in such a case would again be normal with a short range and a mean approximately equal to the true VSD. If incorrectly, a linear model is fit to this response, the VSD would be drastically

FIGURE 3.3:

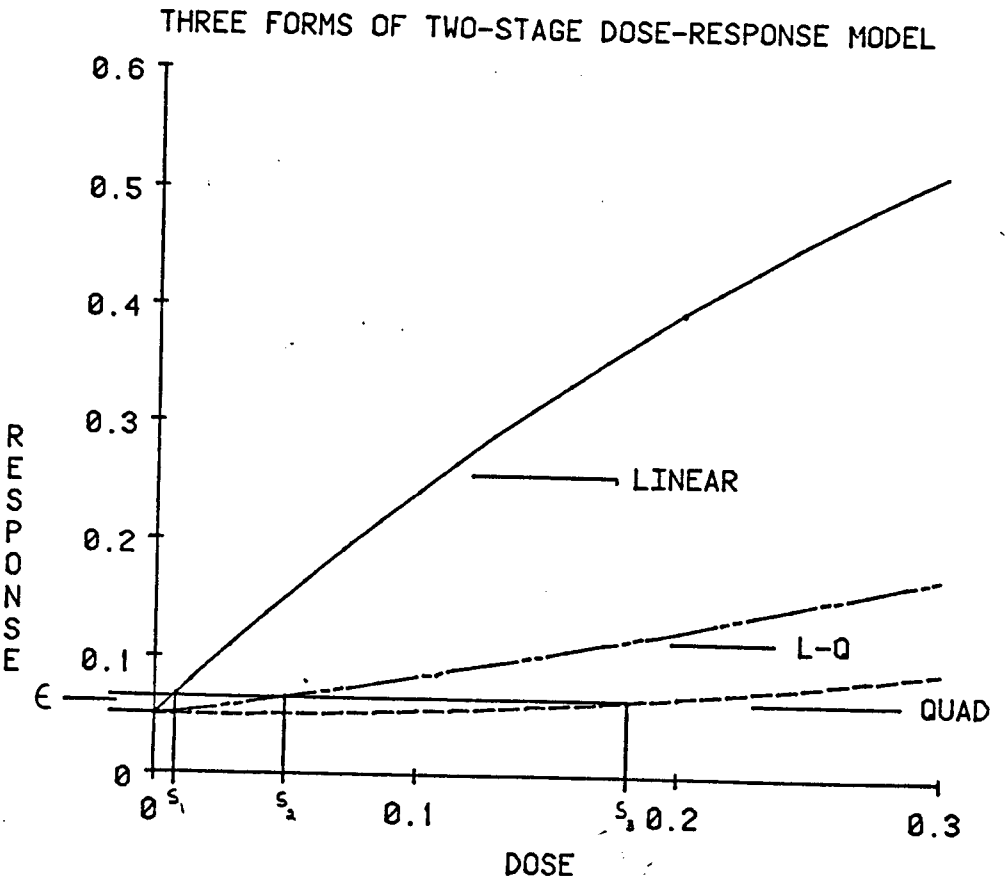
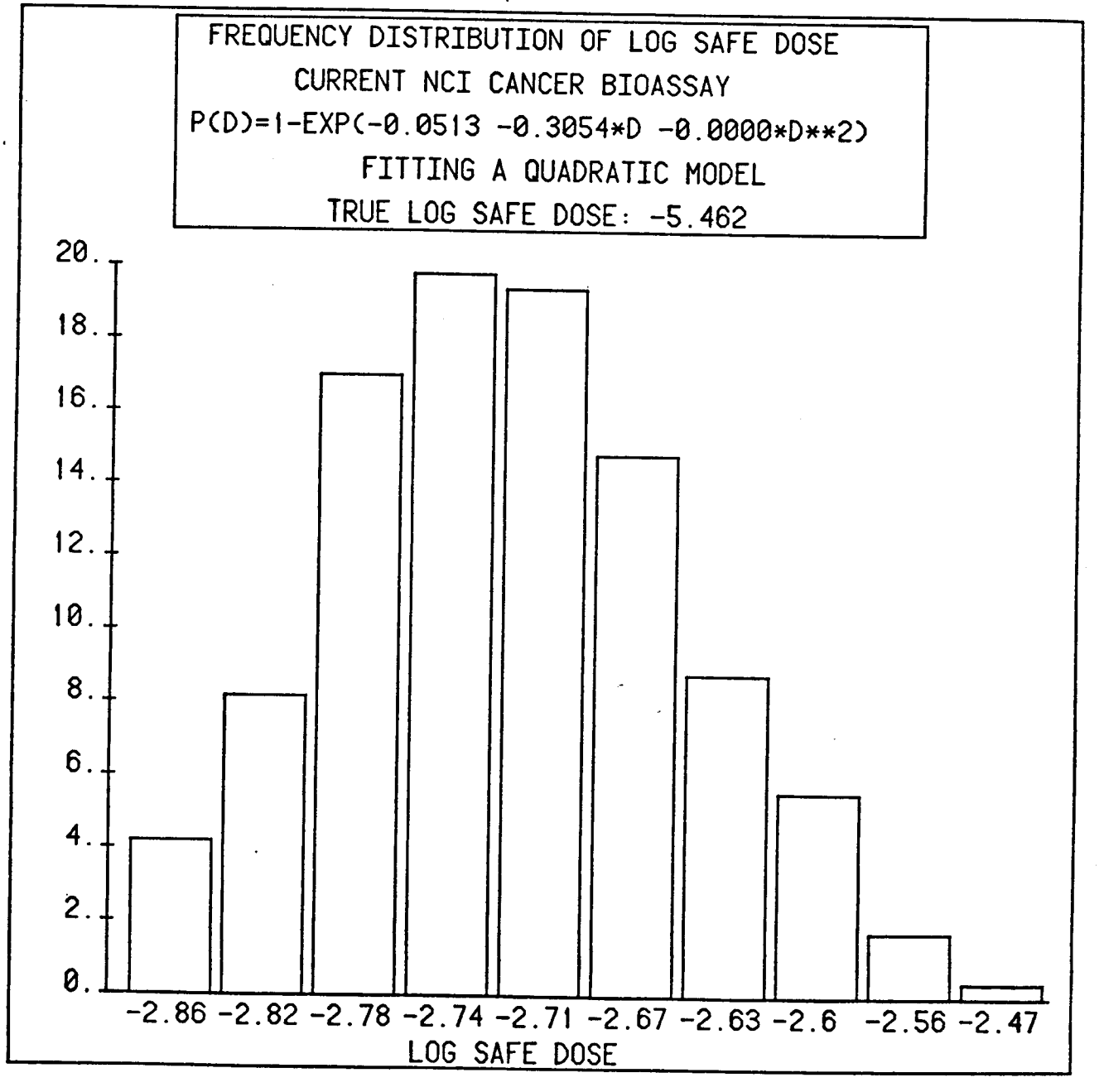

THREE FORMS OF TWO-STAGE DOSE-RESPONSE MODEL
0.6
0.5
0.4
0.3
0.2
0.1
ε
0
RESPONSE
LINEAR
L-Q
QUAD
0
s_1
s_2
0.1
s_3
0.2
0.3
DOSE

FIGURE 3.4:


FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3054*D -0.0000*D**2)
FITTING A QUADRATIC MODEL
TRUE LOG SAFE DOSE: -5.462
20.
18.
16.
14.
12.
10.
8.
6.
4.
2.
0.
-2.86 -2.82 -2.78 -2.74 -2.71 -2.67 -2.63 -2.6 -2.56 -2.47
LOG SAFE DOSE

underestimated thus overestimating the risk of exposure to an agent which has this underlying dose-response.

In general, when dealing solely with linear and quadratic models, it is a win all or lose all situation. If the proper assumptions are made and the correct model is fit, the estimates of the VSD are close and generally reliable. If however, the assumptions are incorrect and the wrong model is fit, the estimate of the VSD is likely to be 2 to 3 orders of magnitude off.

Similar comments apply to the situation in which the underlying model is linear-quadratic and either a linear or a quadratic model is fit. The VSD will be 1 to 2 orders of magnitude too small if the linear model is used (Figure 3.5) and if a quadratic model is used, the VSD can be overestimated by 1 to 2 orders of magnitude (Figure 3.6). Thus, when the underlying dose-response follows a linear-quadratic model, the errors incurred by fitting a linear or quadratic model are smaller than those mentioned previously for underlying linear and quadratic models but are still large enough to warrant the use of additional information to decide the form of the model prior to fitting.

Instead of choosing to fit a linear or a quadratic model, it is possible to let the experimental outcome decide the form of the model. If a linear-quadratic model is fit, it is possible to achieve any of the three models; linear ($\alpha_0 \geq 0$, $\alpha_1 > 0$, $\alpha_2 = 0$), quadratic ($\alpha_0 \geq 0$, $\alpha_1 = 0$, $\alpha_2 > 0$) or linear-quadratic ($\alpha_0 \geq 0$, $\alpha_1 > 0$, $\alpha_2 > 0$). The form of the model being fit to response from an animal experiment is decided by maximizing the likelihood over all possible models. The effect of fitting this type of model on dose-response outcome from any of the various underlying models can be investigated. The assumption

FIGURE 3.5:

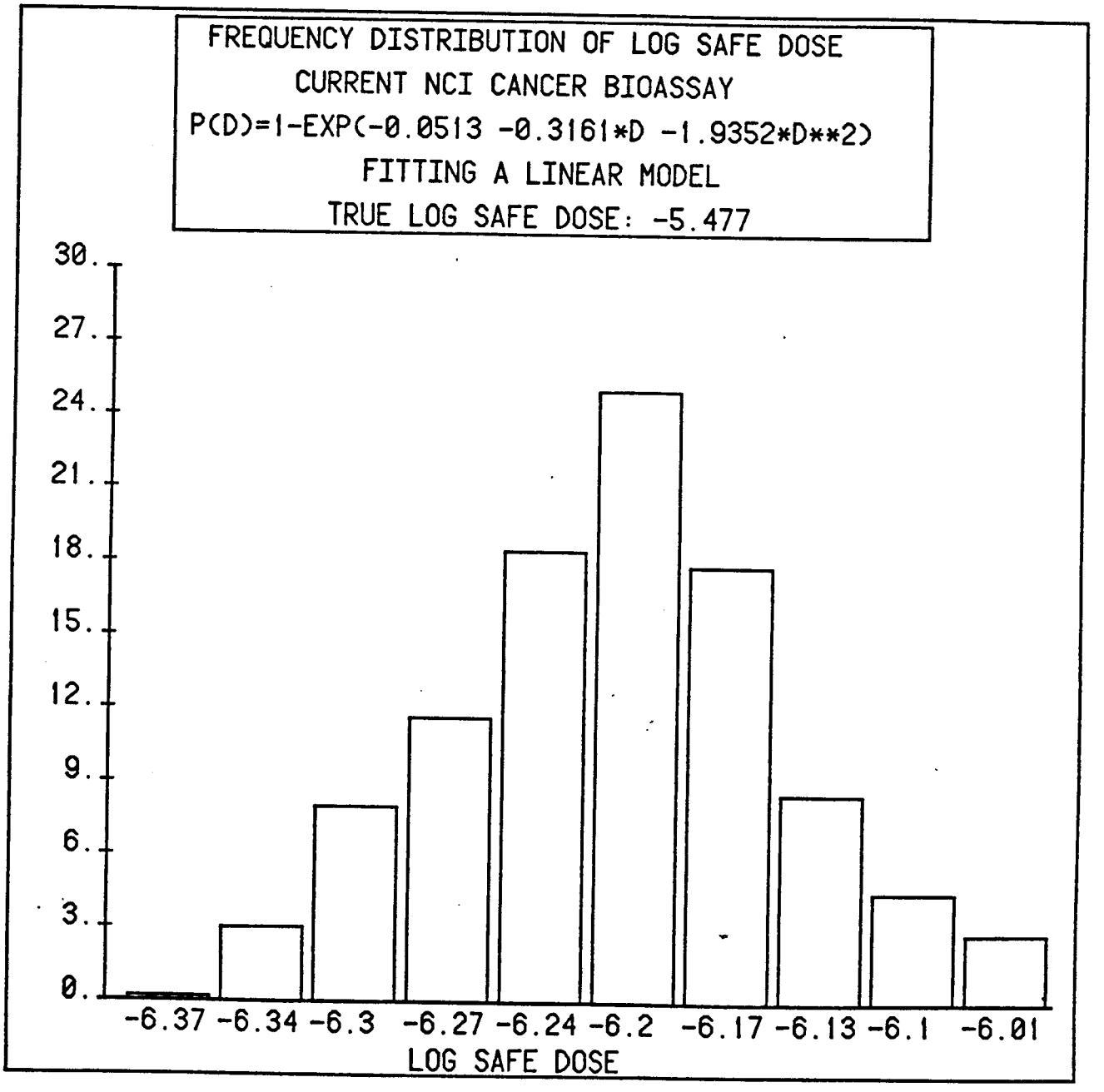
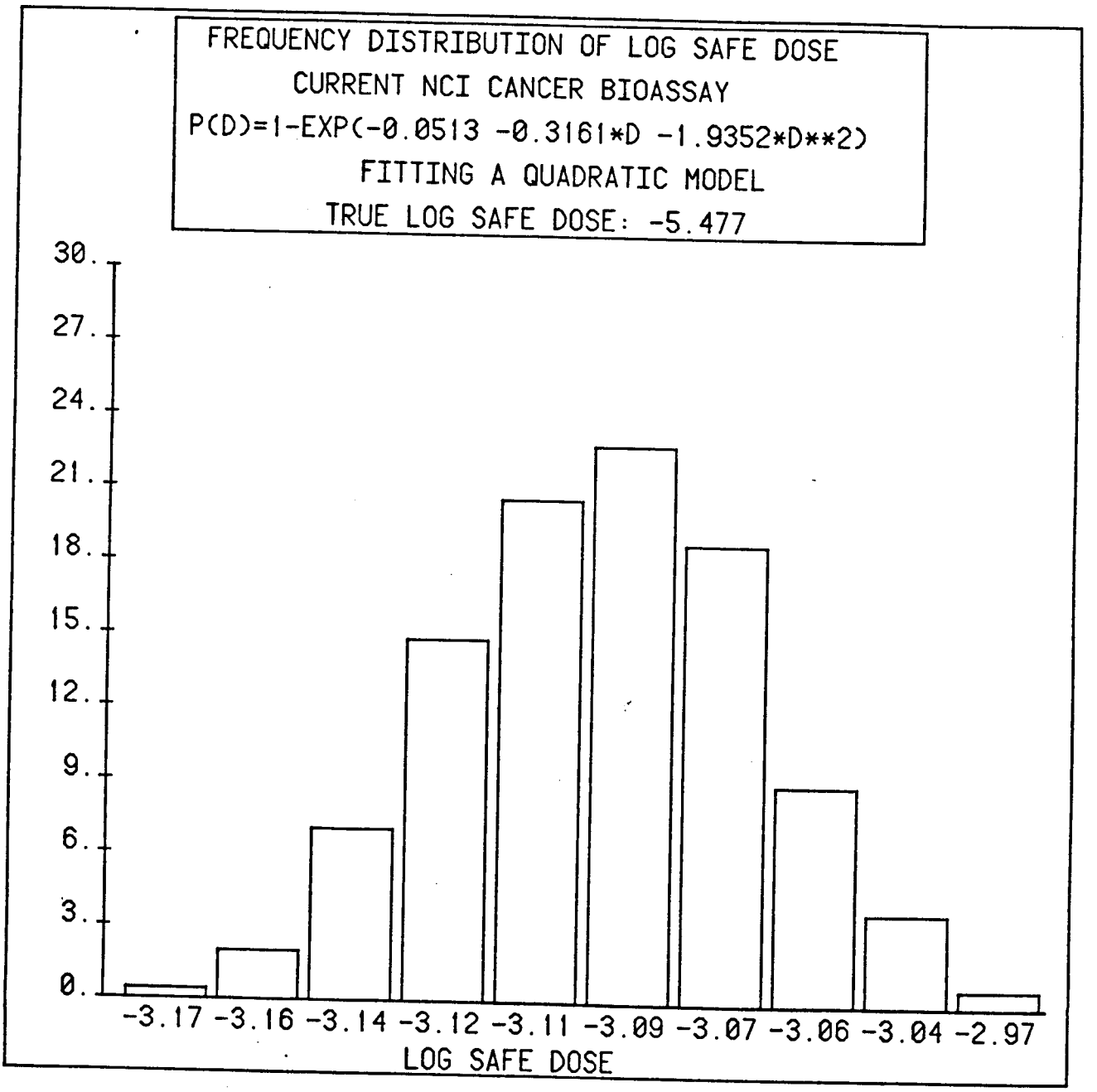

FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3161*D -1.9352*D**2)
FITTING A LINEAR MODEL
TRUE LOG SAFE DOSE: -5.477
30.
27.
24.
21.
18.
15.
12.
9.
6.
3.
0.
-6.37 -6.34 -6.3 -6.27 -6.24 -6.2 -6.17 -6.13 -6.1 -6.01
LOG SAFE DOSE

FIGURE 3.6:


FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3161*D -1.9352*D**2)
FITTING A QUADRATIC MODEL
TRUE LOG SAFE DOSE: -5.477
30.
27.
24.
21.
18.
15.
12.
9.
6.
3.
0.
-3.17 -3.16 -3.14 -3.12 -3.11 -3.09 -3.07 -3.06 -3.04 -2.97
LOG SAFE DOSE

here is that by using this method the chances of fitting the wrong model will be reduced since, the model being fit has the highest likelihood of being correct.

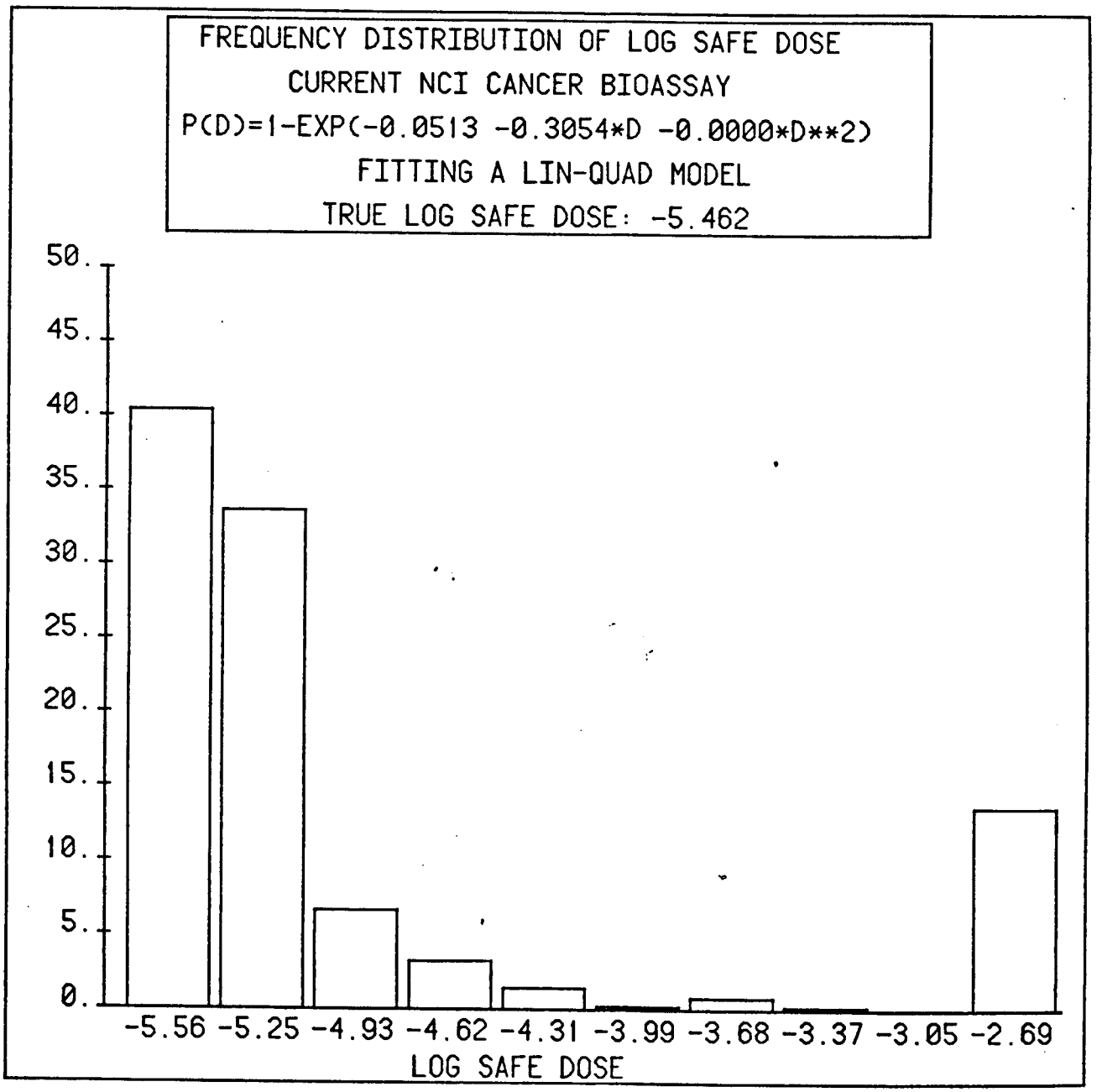
Again assume that, for some agent being investigated, the underlying dose-response can be explained by the use of a linear model. If a linear-quadratic model is fit to the results of an experiment on this agent, the distribution of the estimated VSD will be similar to that of Figure 3.7. For the particular case portrayed in Figure 3.7, the NCI design is used with an underlying linear model given by:

$$P(d) = 1 - e^{-.0513 - .3054d} \qquad (3.3.1)$$

and an excess risk of 10^{-6}. Similar histograms are observed for other linear models, designs and excess risks.

The distribution in Figure 3.7 is bimodal and very different from normal. Note that the two modes seem to be at $10^{-5.5}$ and $10^{-2.7}$. By carefully examining the simulation results and locating what mode a particular estimate of the VSD falls into in comparison with the type of model chosen by the likelihood, the cause of the two modes becomes clear. The mode centered around $10^{-5.5}$ results from the choice of fitting a linear model to the simulated outcome. The upper mode centered at $10^{-2.7}$ is the result of the likelihood determining a quadratic model is the best to fit. This type of error occurred 14% of the time. For other models and designs, it can occur from 3% to 60% of the time. Thus, a response which is truly linear can, with high probability fit a quadratic model better than it would fit a linear model. From this figure, it is also possible to argue for the existence of a third mode at $10^{-3.68}$ which results from fitting a linear-quadratic

FIGURE 3.7:


FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3054*D -0.0000*D**2)
FITTING A LIN-QUAD MODEL
TRUE LOG SAFE DOSE: -5.462
50.
45.
40.
35.
30.
25.
20.
15.
10.
5.
0.
-5.56 -5.25 -4.93 -4.62 -4.31 -3.99 -3.68 -3.37 -3.05 -2.69
LOG SAFE DOSE

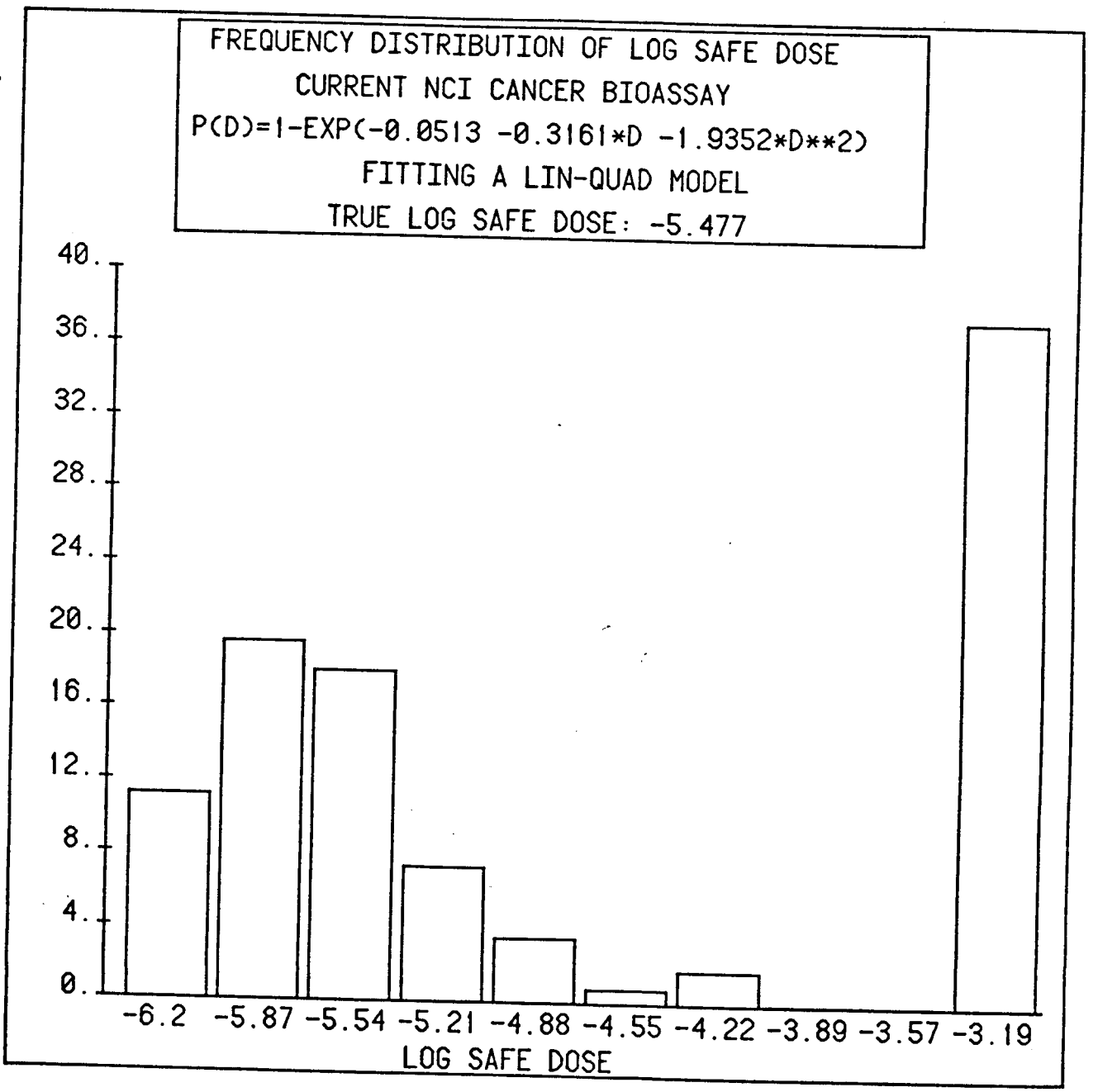
model. In general, the VSD's derived from choosing linear-quadratic models tend to wash into the right tail of the lower mode making this center mode difficult to discern. However, for underlying linear models, these investigations have shown that estimation of a linear-quadratic model using this type of model fitting procedure, can involve 0% to 20% of the fitted models. Thus, it is also seen that response which is linear in nature can look linear-quadratic when the likelihood is used to determine the functional form.

When the underlying model is quadratic, a bimodal distribution is again very likely to be observed. In this case, typically the larger mode would be centered around the true safe dose with a lower mode contributable to estimation of the VSD using linear and linear-quadratic models. The error in model choice for this case is again in the range from 3% to 60%.

The final case deals with fitting a linear-quadratic model to response which is linear-quadratic in nature. Here the bimodality is worse instead of better (a near-normal distribution is expected since the right model is being fit) as shown in Figure 3.8. It is possible to consider the linear-quadratic model as falling between the linear and quadratic models. Responses obtained from this underlying model will tend toward both linear and quadratic with high probability and very seldom fit a linear-quadratic model. Thus, the frequency of fitting an incorrect model increases to range from 5% to 95%.

The implications concerning the distribution of the VSD on the bioassay design are quite important. When fitting a linear or quadratic model, no design modifications will improve the estimation if the incorrect model form is applied. In applying a linear-quadratic model

FIGURE 3.8:


FREQUENCY DISTRIBUTION OF LOG SAFE DOSE
CURRENT NCI CANCER BIOASSAY
P(D)=1-EXP(-0.0513 -0.3161*D -1.9352*D**2)
FITTING A LIN-QUAD MODEL
TRUE LOG SAFE DOSE: -5.477
40.
36.
32.
28.
24.
20.
16.
12.
8.
4.
0.
-6.2
-5.87
-5.54
-5.21
-4.88
-4.55
-4.22
-3.89
-3.57
-3.19
LOG SAFE DOSE

using constrained maximum likelihood, it is seen that response which is non-quadratic can, with positive probability, fit a quadratic model very well, and vice-versa. Therefore, in choosing an optimal design, minimum mean-squared-error, is not the only consideration. A secondary criteria is to minimize the probability of choosing the incorrect model.

The distributions of the VSD for small bioassays and the error in model choice also underscore the need for incorporation of biological justification in the risk assessment. The determination of what model to use can not be simply left to the choice of some mathematical technique but should encompass biological reasoning.

3.4 A Comparison Of Asymptotic vs Small Bioassay VSD Distributions

In studies such as this one, it is of interest to determine how well the asymptotic theory predicts what will occur for smaller numbers. In most cases, it is much simpler to achieve the required outcome by use of asymptotic theory than it is to work with the small sample mathematics. For bioassay design, one question of importance concerns how large the bioassay must be before the asymptotic theory is a good approximation. Also, it is of interest to determine how well confidence bounds based on the asymptotic theory compare to what was realized for the small bioassays.

Two methods will be employed in comparing the asymptotic theory to the results observed for the small bioassay; shape of the distribution and 95% confidence bounds. The situations in which linear or quadratic models are fit will be disregarded. As previously discussed, use of these models involves a hit or miss principle; either the right model form is chosen and good estimates of the VSD result or the wrong

model form is chosen and the estimates of the VSD are consistently 2 to 3 orders of magnitude off. Instead, the more widely accepted linear-quadratic model will be concentrated upon.

The shape of the distribution of the VSD generated by the asymptotic theory can be compared to the frequency histograms observed from the simulation study. Assume the underlying model is linear, specifically:

$$P(d) = 1 - e^{-.0513 - 2.25d} \qquad (3.4.1)$$

Applying the NCI design protocol, by adjusting the mass of the asymptotic distribution, it is possible to overlay this distribution with the frequency distribution in the same plot as shown by Figure 3.9. This plot shows that the asymptotic theory underestimates the mass in the right tail for small samples, completely ignoring the bimodality observed for the simulations. The underlying model (3.4.1) assumes response at the MTD is 90%. As the MTD response gets smaller, the error in the right tail for the asymptotic distribution gets progressively worse. This is portrayed in Figure 3.10 where the underlying model is:

$$P(d) = 1 - e^{-.0513 - .111d} \qquad (3.4.2)$$

a linear model with 15% response at the MTD. The effect of this error on the upper confidence bounds of the VSD will be explored later in this report.

A useful result would be to determine how many animals are required for the small bioassay results to be approximated well by the asymptotic theory. For underlying linear models, if response at the MTD is high (3.4.1), the two distributions converge for as few as 600 animals (200/dose in the NCI protocol). Figure 3.11 illustrates this

FIGURE 3.9:

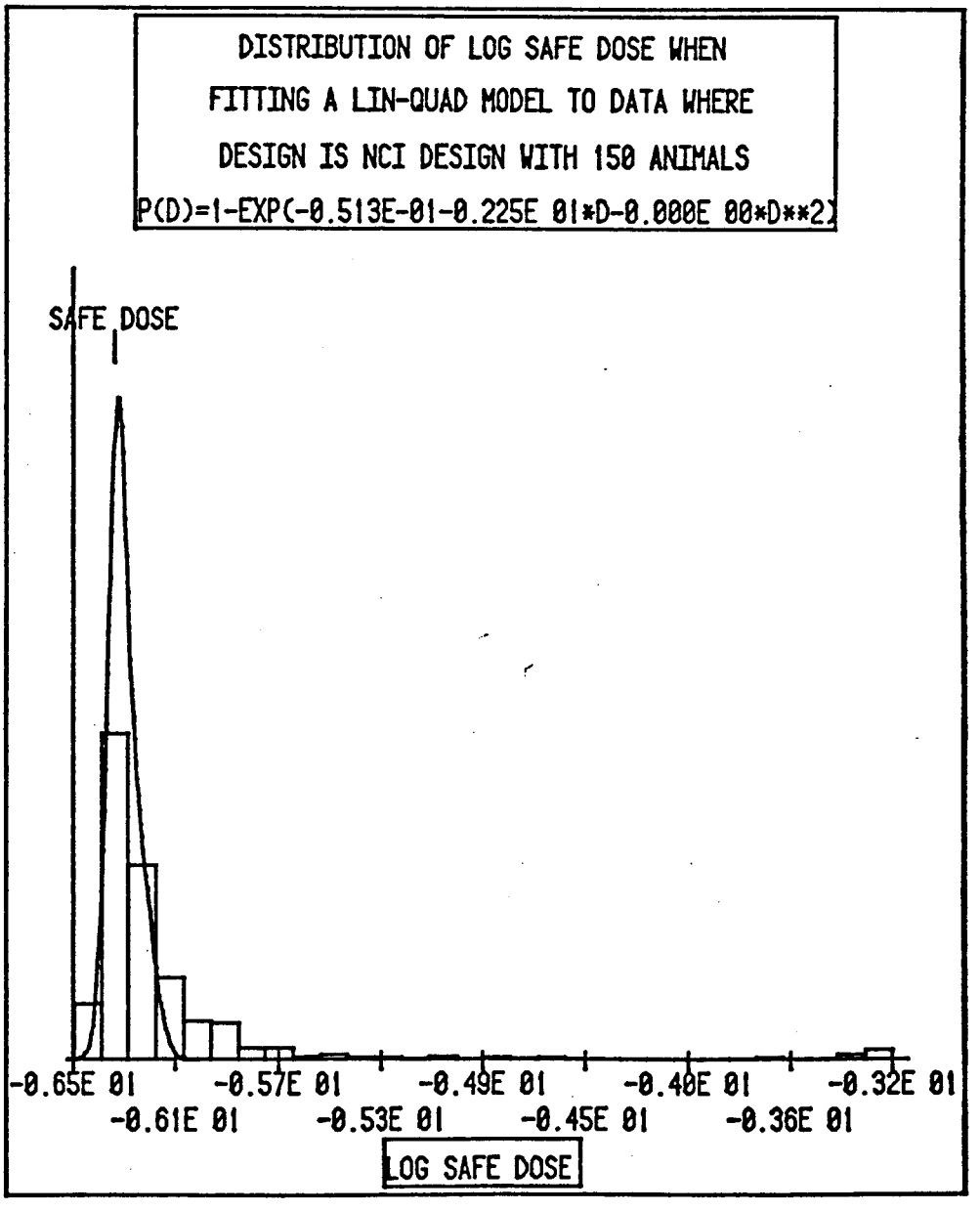

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 150 ANIMALS
P(D)=1-EXP(-0.513E-01-0.225E 01*D-0.000E 00*D**2)
SAFE DOSE
-0.65E 01
-0.61E 01
-0.57E 01
-0.53E 01
-0.49E 01
-0.45E 01
-0.40E 01
-0.36E 01
-0.32E 01
LOG SAFE DOSE

FIGURE 3.10:

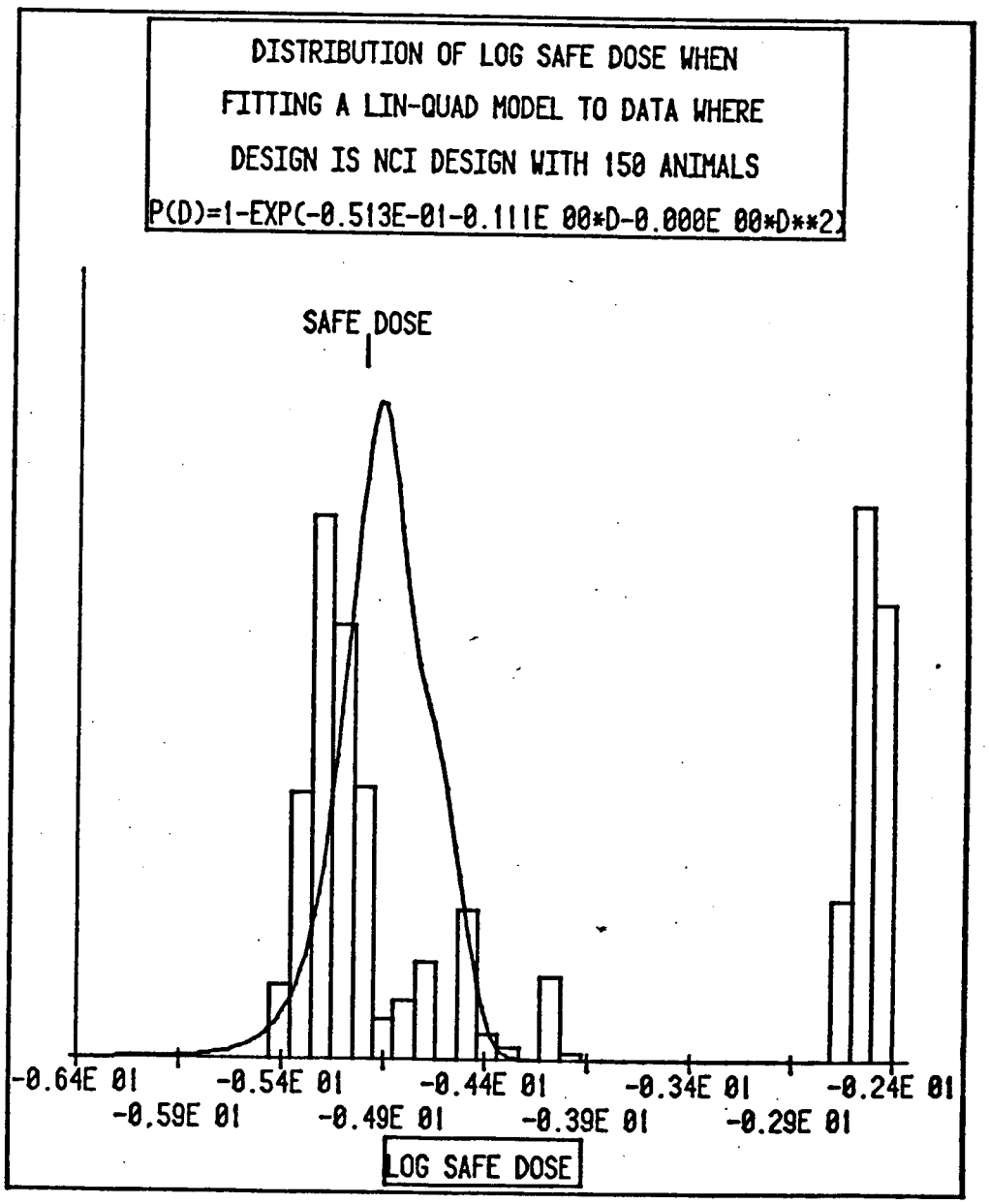
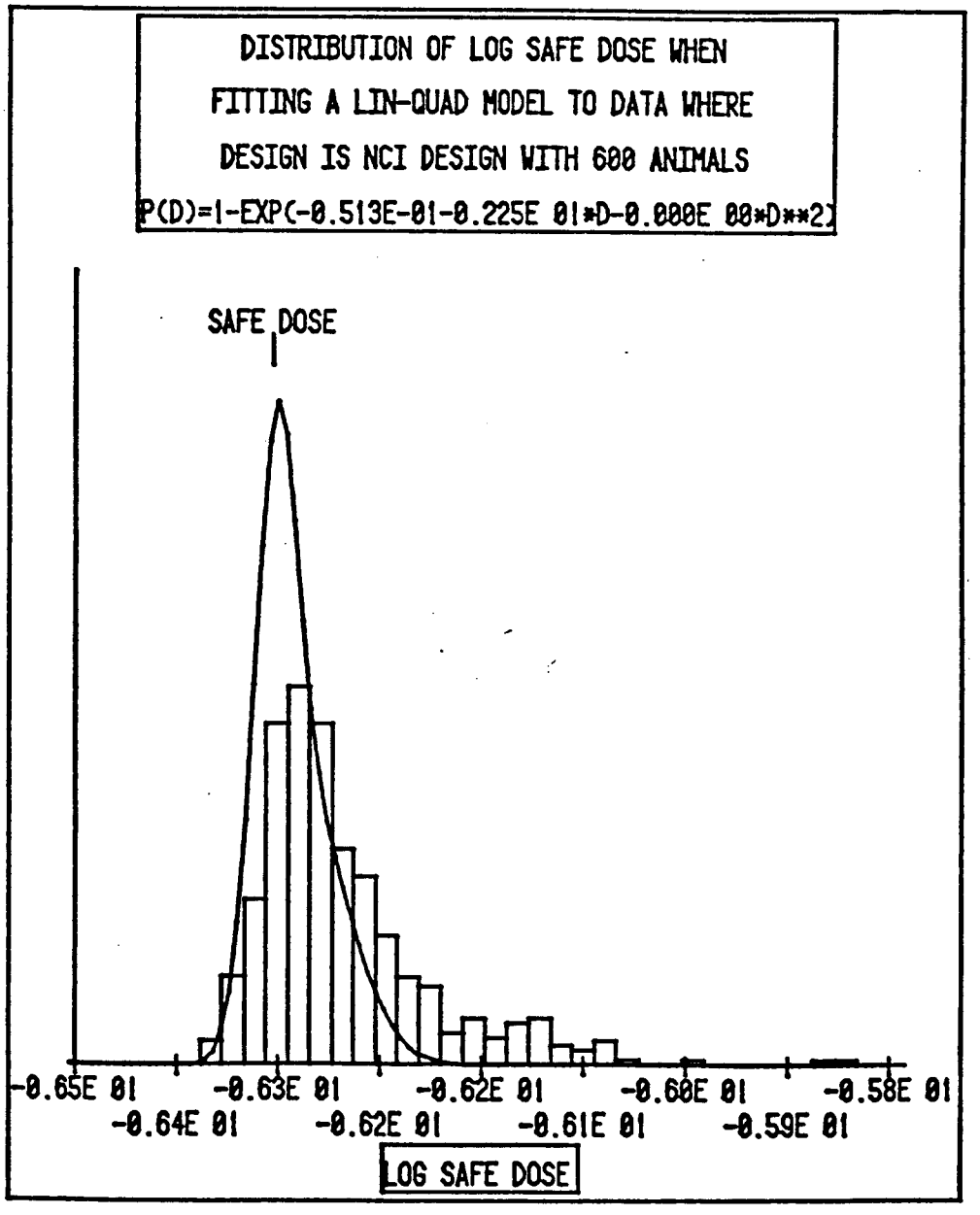

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 150 ANIMALS
P(D)=1-EXP(-0.513E-01-0.111E 00*D-0.000E 00*D**2)
SAFE DOSE
-0.64E 01
-0.59E 01
-0.54E 01
-0.49E 01
-0.44E 01
-0.39E 01
-0.34E 01
-0.29E 01
-0.24E 01
LOG SAFE DOSE

FIGURE 3.11:


DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 600 ANIMALS
P(D)=1-EXP(-0.513E-01-0.225E 01*D-0.000E 00*D**2)
SAFE DOSE
-0.65E 01
-0.64E 01
-0.63E 01
-0.62E 01
-0.62E 01
-0.61E 01
-0.60E 01
-0.59E 01
-0.58E 01
LOG SAFE DOSE

point. However, when the underlying linear model has low response at the MTD, the two distributions don't converge for as many as 10,000 animals. In fact, using the linear model (3.4.2), we see the bimodality of the distribution persists for 10,000 animals as shown in Figure 3.12. For underlying response of this type, the bimodality does not disappear until experiments using over 50,000 animals are employed. However, the two distributions still do not converge for, as shown by Figure 3.13 using 100,000 animals, the right hand tail of the asymptotic distribution is too short to approximate what was observed from the simulations.

When the underlying model is linear-quadratic, the asymptotic theory states that the distribution of the VSD should be normal (2.2.12). However, again due to bimodality, the right hand tail of the distribution using the asymptotic theory dose not correspond to the simulated results. In fact for the typical bioassay (150 animals), the distributions are quite different as shown in Figure 3.14. For this model, the bimodality remains even for large bioassays with as many as 100,000 animals. In Figure 3.15, using a linear-quadratic model with form:

$$P(d) = 1 - e^{-.0513 - .0171d - .0941d^2} \qquad (3.4.3)$$

there is still a trace of bimodality at $10^{-2.4}$ and it is seen that the right-hand tail of the asymptotic distribution is again too short to fully explain the distribution observed from the simulations.

When the underlying model is quadratic, the asymptotic theory predicts the distribution of the VSD as a weighted combination of two normal densities, one of which has a variance so large, it is approximately uniform. The weight is such that this almost uniform density

FIGURE 3.12:

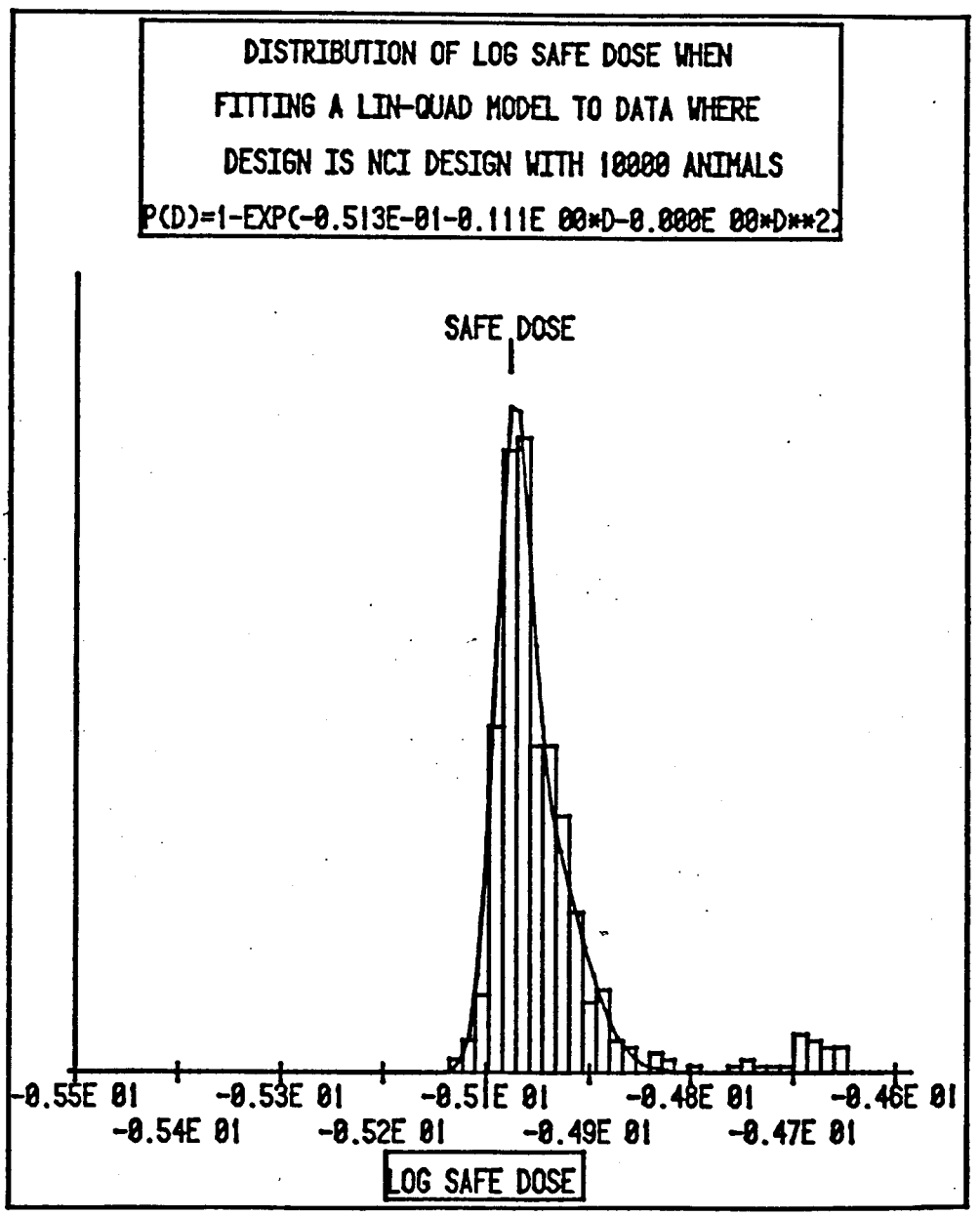

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 10000 ANIMALS
P(D)=1-EXP(-0.513E-01-0.111E 00*D-0.000E 00*D**2)
SAFE DOSE
-0.55E 01
-0.54E 01
-0.53E 01
-0.52E 01
-0.51E 01
-0.49E 01
-0.48E 01
-0.47E 01
-0.46E 01
LOG SAFE DOSE

FIGURE 3.13:

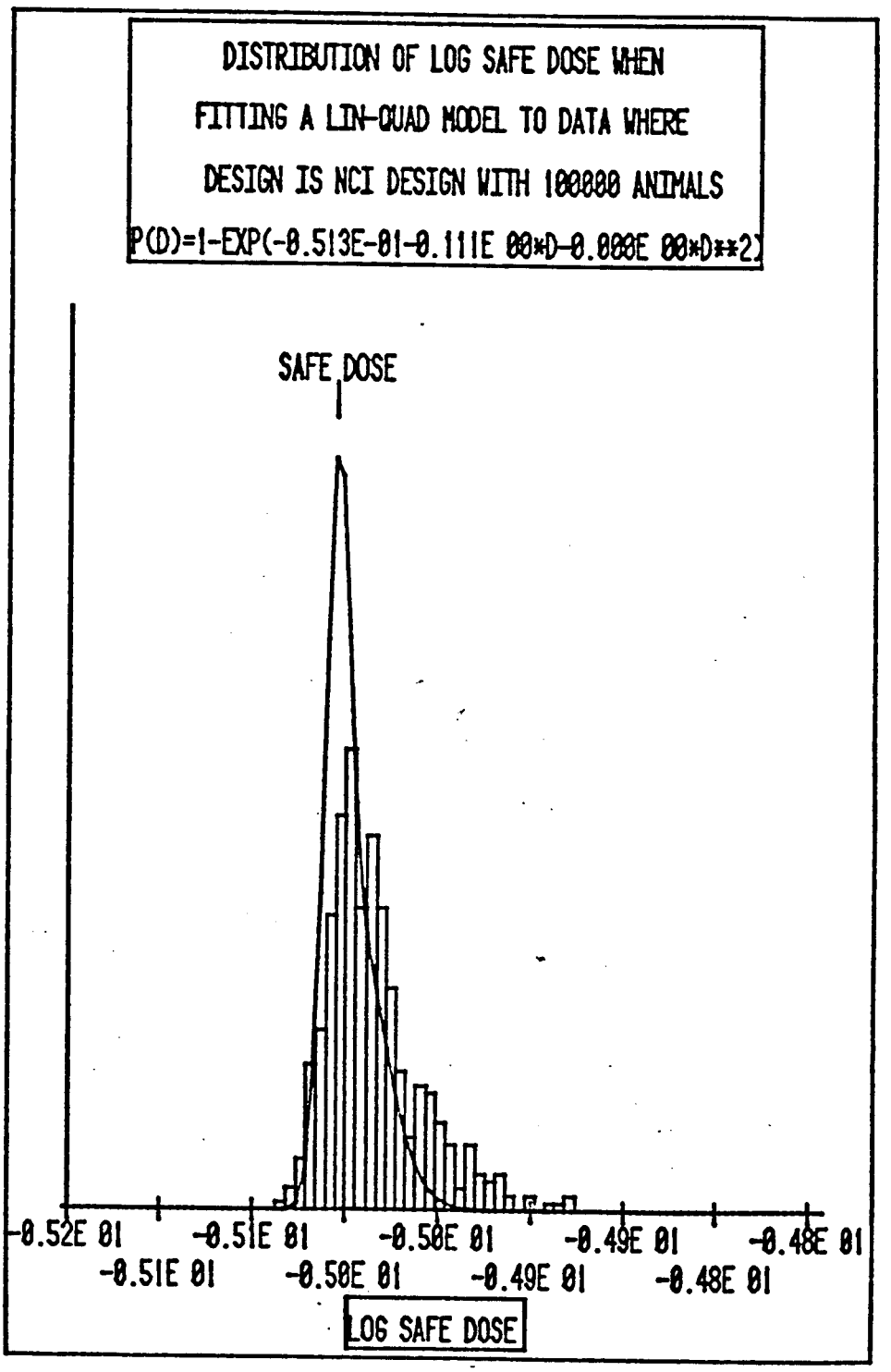

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 100000 ANIMALS
P(D)=1-EXP(-0.513E-01-0.111E 00*D-0.000E 00*D**2)
SAFE DOSE
-0.52E 01
-0.51E 01
-0.50E 01
-0.49E 01
-0.48E 01
-0.51E 01
-0.50E 01
-0.49E 01
-0.48E 01
LOG SAFE DOSE

FIGURE 3.14:

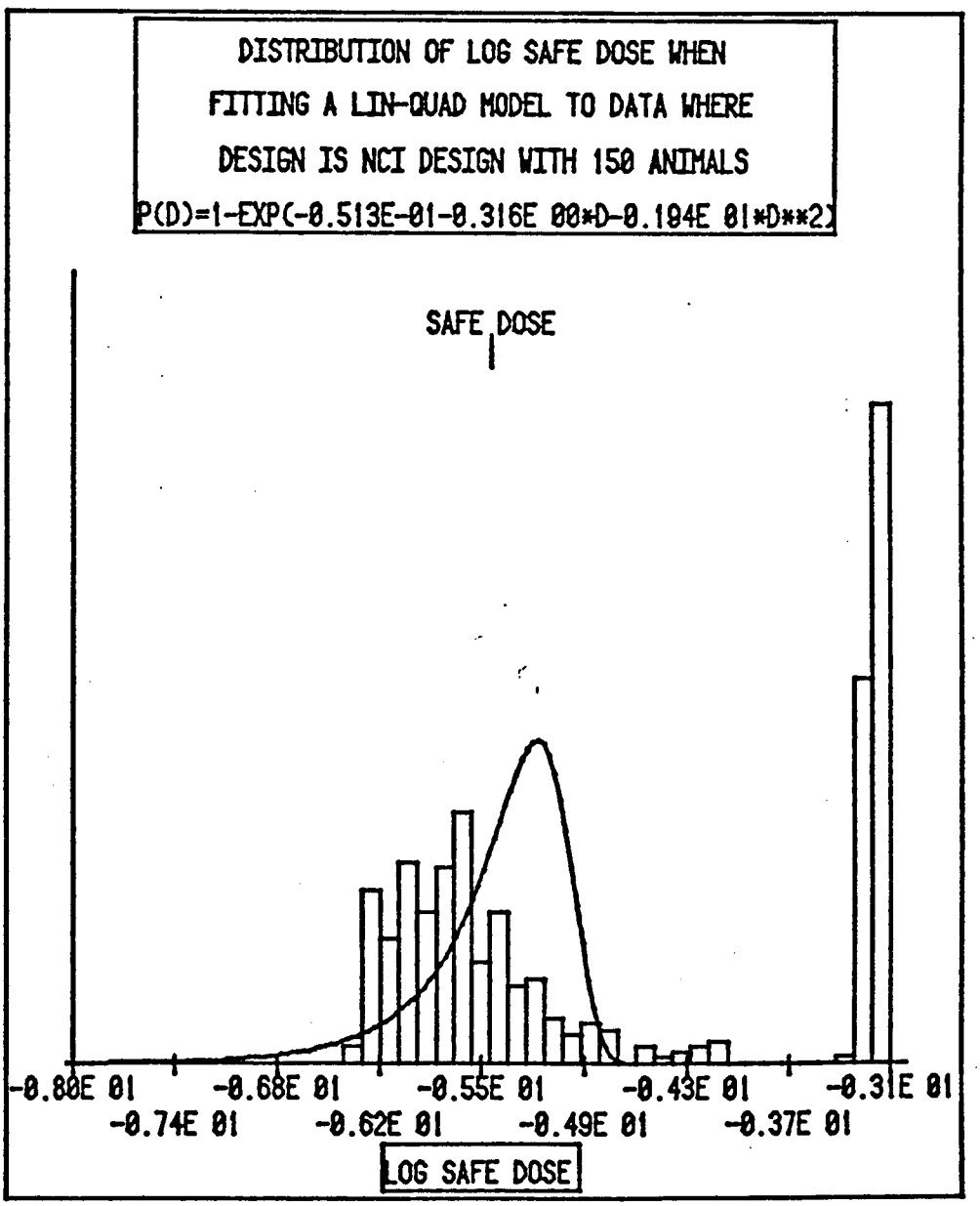
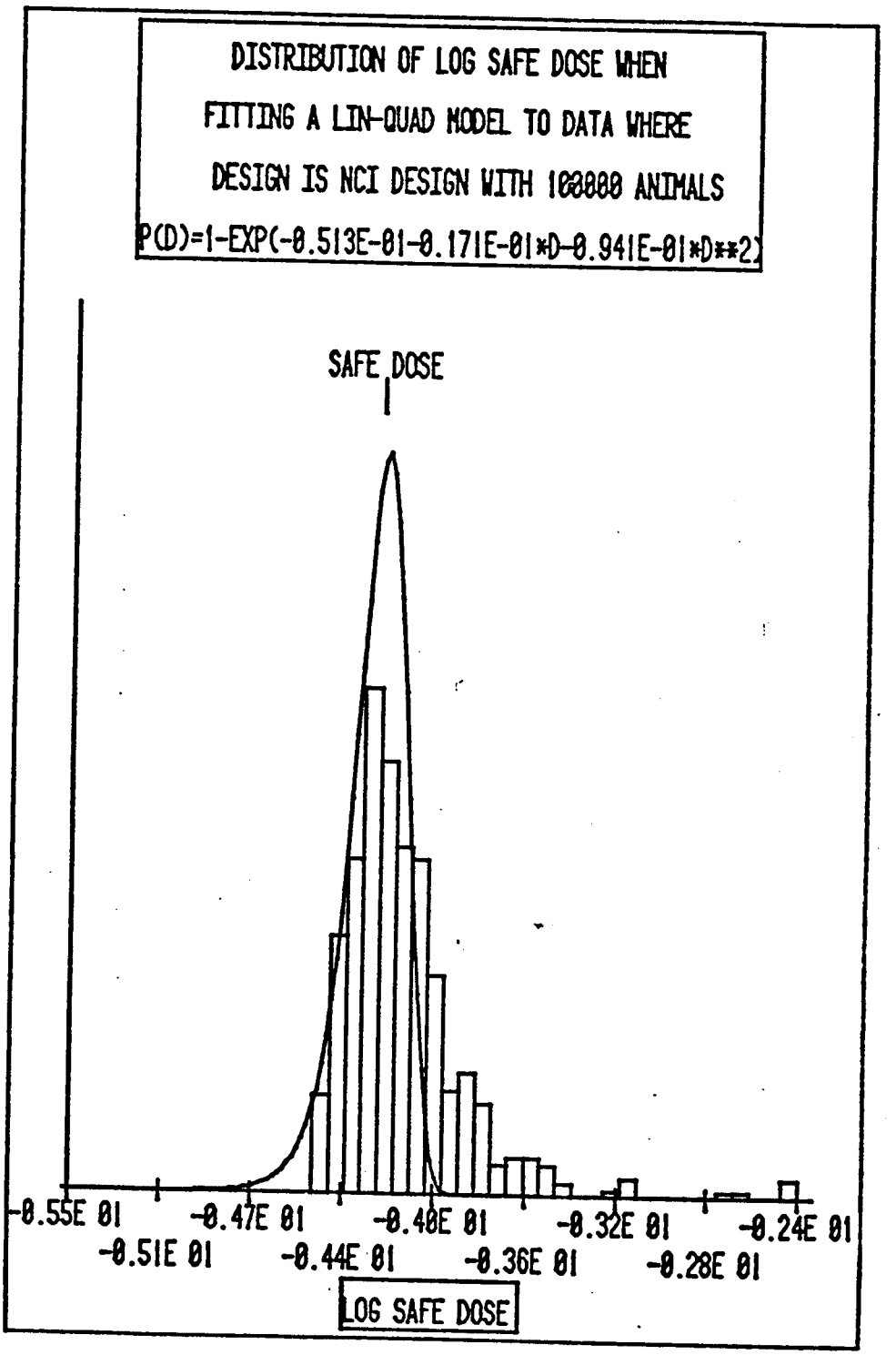

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 150 ANIMALS
P(D)=1-EXP(-0.513E-01-0.316E 00*D-0.194E 01*D**2)
SAFE DOSE
-0.80E 01
-0.74E 01
-0.68E 01
-0.62E 01
-0.55E 01
-0.49E 01
-0.43E 01
-0.37E 01
-0.31E 01
LOG SAFE DOSE

FIGURE 3.15:


DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 100000 ANIMALS
P(D)=1-EXP(-0.513E-01-0.171E-01*D-0.941E-01*D**2)
SAFE DOSE
-0.55E 01
-0.51E 01
-0.47E 01
-0.44E 01
-0.40E 01
-0.36E 01
-0.32E 01
-0.28E 01
-0.24E 01
LOG SAFE DOSE

runs from 0 (possibly negative) to the expected VSD. The resulting density has half of its mass uniformly distributed between zero and the expected VSD and the remaining half normally distributed about the expected VSD. Figure 3.16 illustrates this distribution compared to the distribution observed using Monte Carlo simulations. Again the NCI design protocol with 150 animals is employed and a quadratic underlying model is used of the form:

$$P(d) = 1 - e^{-.0513 - .111d^2} \qquad (3.4.4)$$

For this case, even with the long heavy tail on the left, the asymptotic theory can not account for the bimodality for small samples and does not approximate the frequency distribution very well. However, if an extremely large number of animals is applied in the experiment, the asymptotic theory works well for all cases. Figure 3.17 illustrates this point using the NCI design with 100,000 animals and the quadratic model (3.4.4). It is seen in Figure 3.17 that the left-hand tail of the frequency histogram has flattened out becoming approximately uniform in the range from zero to the expected VSD. This is exactly what the asymptotic theory has predicted.

The second method used to compare the asymptotic theory to the small bioassay results will be to investigate the confidence bounds on the VSD derived from the two. Again, the NCI design protocol for 150 animals is used and any discussion is restricted to the use of an excess risk of 10^{-6}. Note however, that the results presented will be applicable to any excess risk in the range 10^{-8} to 10^{-4}.

Four methods are employed to calculate 95% upper and lower confidence bounds on the VSD estimate, $\hat{s}$. Both upper and lower bounds

FIGURE 3.16:

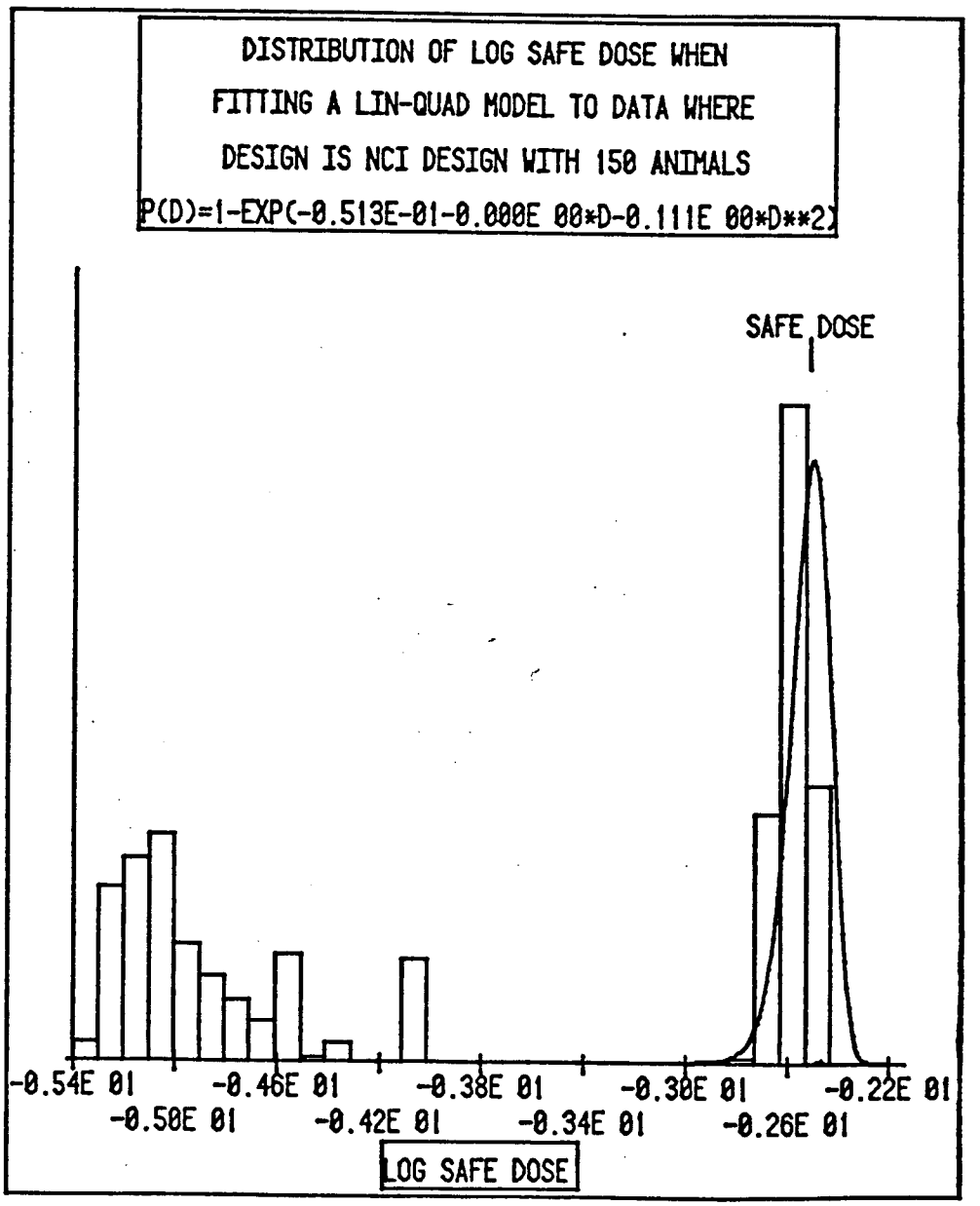
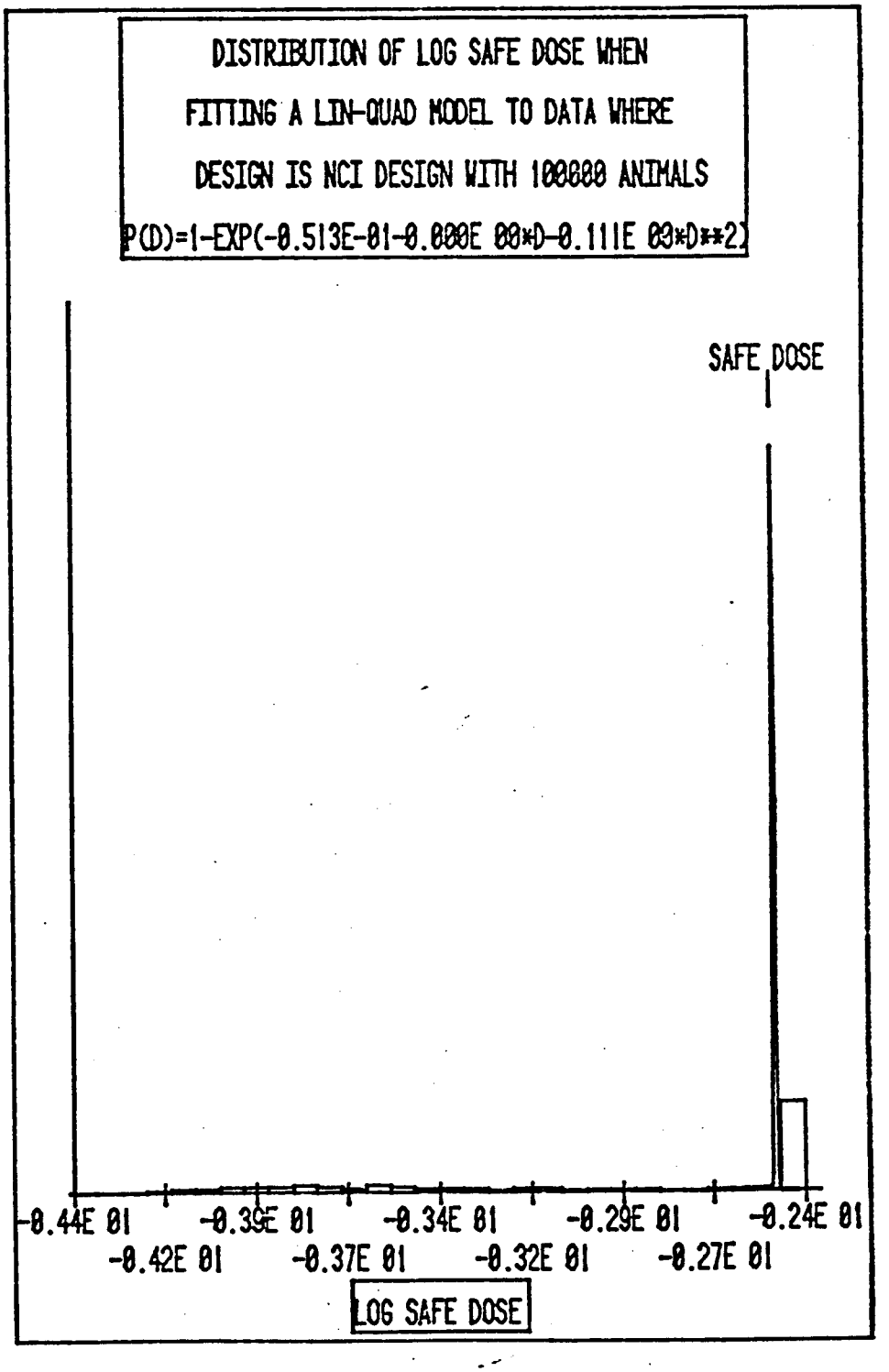

DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 150 ANIMALS
P(D)=1-EXP(-0.513E-01-0.000E 00*D-0.111E 00*D**2)
SAFE DOSE
-0.54E 01
-0.50E 01
-0.46E 01
-0.42E 01
-0.38E 01
-0.34E 01
-0.30E 01
-0.26E 01
-0.22E 01
LOG SAFE DOSE

FIGURE 3.17:


DISTRIBUTION OF LOG SAFE DOSE WHEN
FITTING A LIN-QUAD MODEL TO DATA WHERE
DESIGN IS NCI DESIGN WITH 100000 ANIMALS
P(D)=1-EXP(-0.513E-01-0.000E 00*D-0.111E 03*D**2)
SAFE DOSE
-0.44E 01
-0.42E 01
-0.39E 01
-0.37E 01
-0.34E 01
-0.32E 01
-0.29E 01
-0.27E 01
-0.24E 01
LOG SAFE DOSE

are used since, from a regulatory point of view, the lower bound is important and, in industrial applications, the upper bound can also be of great importance.

Crump (1980) derived a confidence bound on the VSD based on the use of the log-likelihood. Noting that, for very low doses, the dose-response relationship, P, can be approximated by:

$$P(d) = \alpha_1 d \qquad (3.4.5)$$

Crump calculates the $(1 - \alpha)100\%$ lower confidence bound on the safe dose as:

$$s_1 = \varepsilon\ (q_0\ \alpha_1^*)^{-1} \qquad (3.4.6)$$

where $\alpha_1^* > \alpha_1$ is that value of the linear parameter which decreases the log-likelihood L_0, to some value, L_1, such that $2(L_0 - L_1)$ is the cumulative $(1 - 2\alpha)100$ percentage point of the chi-squared distribution with one degree of freedom. Analogously, to find the $(1 - \alpha)100\%$ upper confidence bound on the VSD, (3.4.6) is again used, this time for $\alpha_1^* < \alpha_1$. Note that, in many instances, it is not possible to decrease α_1 enough to change the value of the log-likelihood the necessary amount. In these cases, the upper bound according to Crump's method, is infinite or simply undefined.

The second method concerns the use of the distribution of the VSD for the multi-stage model where the model parameters are unconstrained (UNC). The form of this distribution is given by (2.2.12) for all underlying models. The confidence bound is calculated by inverting the distribution using numerical methods to calculate the variance. It is possible for this distribution to have positive mass for the

VSD estimate less than zero. If this mass is greater than .05, the lower confidence bound will be negative. In such cases, the lower bound is denoted by a "*".

Using similar methods applied to calculate UNC confidence bounds, confidence bounds can be calculated using the asymptotic distribution of the VSD for the true multi-stage model (2.2.1). The distributions in this case are given by (2.2.12) and (2.2.25). Again, due to the assumption that the probability of $\hat{\alpha}_i = 0$ when $\alpha_i > 0$ is zero, it is possible for small bioassays to have positive mass for negative estimates of the VSD.

Finally, confidence bounds can be constructed from the Monte-Carlo simulations by ordering the results from 1000 simulations and choosing the 50th and 950th estimates. These bounds can be considered the true confidence bounds and will be referred to by the label "SIMS".

Table 3.1 gives the upper and lower confidence bounds on the VSD using the four methods previously discussed. The bounds are derived for underlying models given by the 33 models of Table 2.3. The column labeled VSD is the expected value of the virtually safe dose for each model when the excess risk is 10^{-6}.

The lower confidence bounds (LCB) using the UNC method tend to be conservative as compared to the SIMS bounds when they are not zero. Using the asymptotic theory of Chapter II (CON), it is seen that the LCB's are a bit too large, in general, when they are not zero. However, for both of these methods, the LCB's are zero most of the time making them of little use. The CON and UNC methods yield non-zero LCB's more often if the spontaneous response is small, the MTD response is high and the underlying model is close to linear. However, in

TABLE 3.1-A: 95% Lower Confidence Bounds On The VSD

Back-ground	Model Type	MTD Response	VSD	CRUMP	UNC	CON	SIMS
.005	LINEAR	.15	.638E-5	.398E-5	*	.421E-5	.402E-5
		.30	.286E-5	.205E-5	.743E-6	.234E-5	.225E-5
		.50	.146E-5	.113E-5	.657E-6	.126E-5	.114E-5
		.90	.437E-6	.355E-6	.262E-6	.389E-6	.355E-6
	QUADRAT	.15	.253E-2	.400E-5	*	*	.709E-5
		.30	.169E-2	.229E-5	*	*	.400E-5
		.50	.121E-2	.140E-5	*	*	.483E-5
		.90	.661E-3	.621E-6	*	*	.175E-5
	LIN-QUA	.15	.748E-5	.407E-5	*	*	.497E-5
		.30	.361E-5	.212E-5	.397E-6	.397E-6	.234E-5
		.50	.227E-5	.122E-5	.500E-6	.500E-6	.135E-5
		.90	.136E-4	.832E-6	*	*	.172E-5
.050	LINEAR	.15	.946E-5	.473E-5	*	.502E-5	.465E-5
		.30	.345E-5	.230E-5	.138E-6	.262E-5	.246E-5
		.50	.164E-5	.123E-5	.607E-6	.136E-5	.130E-5
		.90	.468E-6	.376E-6	.271E-6	.411E-6	.388E-6
	QUADRAT	.15	.308E-2	.422E-5	*	*	.608E-5
		.30	.186E-2	.237E-5	*	*	.353E-5
		.50	.128E-2	.144E-5	*	*	.363E-5
		.90	.684E-3	.629E-6	*	*	.203E-5
	LIN-QUA	.15	.614E-4	.524E-5	*	*	.497E-5
		.30	.754E-5	.251E-5	*	*	.271E-5
		.50	.347E-5	.138E-5	*	*	.168E-5
		.90	.333E-5	.683E-6	*	*	.128E-5
.200	LINEAR	.30	.936E-5	.384E-5	*	.262E-5	.332E-5
		.50	.266E-5	.174E-5	*	.189E-5	.159E-5
		.90	.601E-6	.470E-6	.307E-6	.505E-6	.458E-6
	QUADRAT	.30	.306E-2	.515E-6	*	*	.369E-5
		.50	.163E-2	.306E-5	*	*	.267E-5
		.90	.775E-3	.151E-5	*	*	.149E-5
	LIN-QUA	.30	.100E-4	.385E-5	*	*	.306E-5
		.50	.522E-5	.185E-5	*	*	.172E-5
		.90	.563E-5	.795E-6	*	*	.137E-5

* indicates lower confidence bound would be negative

TABLE 3.1-B: 95% Upper Confidence Bounds On The VSD

Back-ground	Model Type	MTD Response	VSD	CRUMP	UNC	CON	SIMS
.005	LINEAR	.15	.638E-5	#	.133E-4	.133E-4	.542E-2
		.30	.286E-5	#	.494E-5	.494E-5	.301E-2
		.50	.146E-5	.964E-5	.225E-5	.225E-5	.358E-4
		.90	.437E-6	.120E-5	.608E-6	.608E-6	.404E-5
	QUADRAT	.15	.253E-2	#	.416E+0	.285E-2	.692E-2
		.30	.169E-2	#	.273E+0	.184E-2	.400E-2
		.50	.121E-2	#	.200E+0	.129E-2	.273E-2
		.90	.661E-3	#	.136E+0	.694E-3	.145E-2
	LIN-QUA	.15	.748E-5	#	.167E-4	.167E-4	.582E-2
		.30	.361E-5	#	.679E-5	.679E-5	.337E-2
		.50	.227E-5	#	.402E-5	.402E-5	.452E-3
		.90	.136E-4	#	.130E-3	.130E-3	.140E-2
.050	LINEAR	.15	.946E-5	#	.275E-4	.275E-4	.615E-2
		.30	.345E-5	#	.672E-5	.672E-5	.363E-2
		.50	.164E-5	.819E-4	.266E-5	.266E-5	.211E-2
		.90	.468E-6	.138E-5	.659E-6	.659E-6	.268E-5
	QUADRAT	.15	.308E-2	#	.847E+0	.378E-2	.636E-2
		.30	.186E-2	#	.378E+0	.208E-2	.476E-2
		.50	.128E-2	#	.241E+0	.138E-2	.307E-2
		.90	.684E-3	#	.144E+0	.725E-3	.154E-2
	LIN-QUA	.15	.614E-4	#	.748E-3	.748E-3	.623E-2
		.30	.754E-5	#	.217E-4	.217E-4	.403E-2
		.50	.347E-5	#	.749E-5	.749E-5	.263E-2
		.90	.333E-5	#	.105E-4	.105E-4	.144E-2
.200	LINEAR	.30	.936E-5	#	.364E-4	.364E-4	.473E-2
		.50	.266E-5	#	.543E-5	.543E-5	.364E-2
		.90	.601E-6	.254E-5	.890E-6	.892E-6	.487E-5
	QUADRAT	.30	.306E-2	#	.136E+1	.407E-2	.475E-2
		.50	.163E-2	#	.447E+0	.186E-2	.401E-2
		.90	.775E-3	#	.174E+0	.837E-3	.173E-2
	LIN-QUA	.30	.100E-4	#	.410E-4	.410E-4	.472E-2
		.50	.522E-5	#	.151E-4	.151E-4	.394E-2
		.90	.563E-5	#	.247E-4	.247E-4	.175E-2

# indicates it is impossible to achieve an upper bound

general these methods give poor estimates of the LCB's.

The methodology developed by Crump seems to work very well for calculating LCB's. The use of the likelihood is similar to Monte-Carlo simulation in the estimates of the LCB's since the assumption of a chi-squared distribution for the difference of the two log-likelihoods, L_0 and L_1, is appropriate for 150 animals. The Crump estimates tend to be most conservative when the MTD response is small and the spontaneous background response is small and the underlying model is quadratic. However, in general, the LCB given by Crump's method seems acceptable for small bioassay results. For upper confidence bounds (UCB), the method proposed by Crump is useless since, in virtually every case, the upper bound is undefined (infinite). The UCB's calculated using the CON method tend to be far too small (sometimes by two orders of magnitude) when compared to what was observed for the simulation study. Thus, for both upper and lower confidence bounds, the methodology derived in Chapter II is useless for 150 animals.

The UNC confidence bounds are worse than those for the CON method. When the underlying model is linear or linear-quadratic, this method underestimates the 95% UCB. For quadratic models, it overestimates the UCB.

In this section, the asymptotic theory of Chapter II was compared to what was observed using Monte Carlo simulation for small bioassays. For the typical bioassay of 150 animals, the distribution of the VSD resulting from use of the asymptotic theory differed drastically from what was observed in the simulations. When response at the MTD is high, the asymptotic theory approximates the distribution of the VSD well for moderate numbers of animals (600). However, for most underlying models,

a large number of animals (100,000) are required to get a good approximation of the distribution of the VSD from the asymptotic theory.

Four methods were used to determine confidence bounds on the VSD. The bounds calculated using the results of the simulation study could be considered the true confidence bounds. The UNC bounds based on the asymptotic theory ignoring the constrained nature of the parameters in the multi-stage model failed to produce good estimates of the upper and lower confidence bounds of the VSD. The failure of the CON method, developed in Chapter II, to correctly estimate the confidence bounds on the VSD was due to a failure of one of the assumptions used to develop this theory. Asymptotically, it was proven that the probability of the estimate of $\hat{\alpha}$ to be zero when the expected value was non-zero was zero. For small bioassays involving only 150 animals, this assumption is incorrect as shown by the bimodality of the estimated VSD. Inclusion of this case may lead to better approximations using the asymptotic theory.

Finally, confidence bounds developed by Crump gave excellent estimates of the lower confidence bound on the VSD. Here, the lower bound includes a positive probability of $\hat{\alpha}$ being greater than zero for all underlying models. This, and the validity of the chi-squared distribution for the difference in the log-likelihood combine to mimic the results for lower confidence bounds observed in the simulations. This method fails as an upper bound, since it is impossible in many cases to decrease the value of $\hat{\alpha}_1$ enough to produce the required difference in the log-likelihoods. Possible modifications of this procedure utilizing α_2 as well as α_1 may remedy this failure.

3.5 Small Bioassay Optimal Designs

The purpose of this research is to investigate the effect of changing the design parameters, dose and animal allocation, on the estimation of the VSD. The optimal design will be that design, ω, which minimizes the mean-squared-error of the maximum likelihood estimate of the VSD based on a random sample of replicated experiments using Monte Carlo simulation. Note that the study is censored to include only experiments which have shown significant dose-response increase based on the Armitage linear trend test. Discussion of uncensored results is left for Chapter IV.

Again, the discussion concentrates on using the 33 models of Table 2.3 as portraying the range of possible dose-response observed in nature. Optimal designs for fitting linear, quadratic and linear-quadratic models to response which is linear, quadratic and linear-quadratic in nature are discussed.

Suppose an experiment has been conducted to determine the toxicity of some environmental chemical and the analyst has decided to fit a linear model to the response observed. If the underlying model for dose-response was linear, regardless of what design had been used to conduct the experiment (provided it was practical), the accuracy of the estimated VSD would remain unchanged. In this situation, the probability of being more than one order of magnitude off is approximately zero and does not change with design changes. The design which optimizes the estimation (only slightly) would use only a control dose with about 20% of the animals and the remaining animals at the MTD.

Given the same experimental situation with the underlying model quadratic or linear-quadratic, no optimal design exists. The im-

proper assumption of fitting a linear model has so biased the estimate of the VSD, that the magnitude of the mean-squared-error makes it impossible to optimize. In this situation, the probability of being within one order of magnitude of the expected VSD is very small.

Similar comments apply to the use of a quadratic model to approximate dose-response from an animal bioassay. When the underlying response follows a quadratic structure, the optimal design will have two groups, control and MTD, with approximately 30% of the animals at control. For underlying linear or linear-quadratic models, no design changes will help to correct the bias incurred from use of incorrect assumptions used in deciding the form of the dose-response from the experiment.

The remainder of the discussion will concentrate on the use of the linear-quadratic model in approximating dose-response behavior. For this model, design changes can have marked effect on the estimated mean-squared-error. To locate optimal designs, a directed grid search over the parameters of interest was used to find the dose and animal allocations. A directed grid search in this context involves the use of smaller and smaller grids until the design is within a pre-defined neighborhood of the global optimal design. The increment for dose in the finest grids was .025 and 10 animals per group was the smallest increment used to allocate animals. For each point on a grid (design), 1000 simulated sets of quantal response data were used to calculate the sample mean-squared-error. Table 3.2 illustrates a typical grid, in this case, for the underlying linear-quadratic model:

TABLE 3.2: Sample Grid Of Mean-Squared-Errors* Used To Find Small Sample Bioassay Designs

#of Control Animals	#of MTD Animals	Middle-Dose (d_1) .1	.2	.3	.4	.5
10	10	.11944	.09926	.09181	.08480	.09590
10	30	.14906	.14014	.12629	.10854	.10682
10	50	.16442	.14733	.13053	.11012	.10994
10	70	.15140	.13910	.12814	.11429	.12048
30	10	.11760	.08998	.07647	.09300	.09844
30	30	.15979	.12423	.10762	.08582	.09378
30	50	.15028	.13169	.11141	.11227	.11038
30	70	.15746	.12938	.12238	.11414	.12384
50	10	.11420	.08460	.07786	.08219	.09462
50	30	.14224	.11993	.10229	.10905	.10512
50	50	.17170	.13682	.11895	.11369	.11494
50	70	.15220	.14099	.13115	.11792	.11914
70	10	.13172	.08577	.09377	.08582	.10010
70	30	.15843	.13900	.11536	.10144	.11076
70	50	.15240	.14737	.13838	.12087	.13127
70	70	.19293	.15900	.13833	.15747	.16808

* entries are $MSE(\hat{s}) \cdot 10^5$

$$P(d)= 1 - e^{-.0513-.140d-.166d^2} .$$

Several practical assumptions were made concerning certain fixed parameters in the design. In all cases it was assumed that the optimal design must contain a control group (d_0=0). This requirement is practical for more reasons than just minimizing the mean-squared-error of the estimate of the VSD (i.e., biological validity of the experiment, statistical testing, ...). It was also assumed for the grid search, that the upper dose, d_2, was fixed at the MTD. This assumption will be tested later in this section and discussed in detail in Chapter IV. Finally, it was assumed that no group (control or dosed) use less than 10 animals. This boundary condition was seldom encountered for the 33 underlying models studied.

Thus, there are only three parameters to optimize over; d_1, γ_0 and γ_2 (note that $\gamma_1=1-\gamma_0-\gamma_2$). From Table 3.2 it is seen that the manifold describing the mean-squared-error seems smooth and well behaved. This same behavior was observed in every grid and the choice of optimal design could be narrowed down to one or two sections of each grid, eventually locating the optimal design.

The SO designs for the 33 models of Table 2.3 are shown in Table 3.3 when the linear-quadratic model is used to approximate the observed dose-response. The models are portrayed in the same manner used in Table 2.3. The column labeled MSE($\hat{s}$) is the value of the mean-squared-error of the VSD calculated from the simulations for that design using formula (3.2.1).

The SO designs for agents in which the underlying dose-response behavior is linear or linear-quadratic seem to be constant over all

TABLE 3.3: Small Bioassay Optimal Designs For An Added Risk Of 10^{-6}

Back-ground	Model Type	MTD Response	d_1	# At Control	# At d_1	# At MTD	MSE($\hat{s}$)
.005	LINEAR	.15	.175	10	130	10	.458E-6
		.30	.300	30	90	30	.459E-7
		.50	.300	20	110	20	.140E-8
		.90	.200	10	75	65	.189E-13
	QUADRAT	.15	.075	70	10	70	.536E-6
		.30	.050	70	10	70	.200E-6
		.50	.050	70	10	70	.948E-7
		.90	.050	70	10	70	.532E-7
	LIN-QUA	.15	.300	10	120	20	.533E-6
		.30	.200	10	130	10	.133E-6
		.50	.300	20	100	30	.189E-7
		.90	.125	10	110	30	.132E-6
.050	LINEAR	.15	.300	10	130	10	.867E-6
		.30	.300	30	110	10	.289E-6
		.50	.300	70	70	10	.353E-7
		.90	.200	20	70	60	.468E-13
	QUADRAT	.15	.100	70	10	70	.367E-5
		.30	.075	70	10	70	.131E-5
		.50	.050	70	10	70	.635E-6
		.90	.150	60	80	10	.171E-6
	LIN-QUA	.15	.300	10	130	10	.111E-5
		.30	.300	30	110	10	.765E-6
		.50	.300	30	110	10	.231E-6
		.90	.200	30	90	30	.113E-6
.200	LINEAR	.30	.300	50	90	10	.531E-6
		.50	.300	50	80	20	.231E-6
		.90	.200	50	50	50	.326E-8
	QUADRAT	.30	.100	70	10	70	.466E-5
		.50	.100	50	50	50	.122E-5
		.90	.275	70	70	10	.243E-6
	LIN-QUA	.30	.300	30	110	10	.541E-6
		.50	.400	50	80	20	.464E-6
		.90	.325	30	50	70	.227E-6

such parameterizations. Generally, the middle dose, d_1, is approximately 30% of the MTD and receives between 70 and 100 of the animals. The remaining animals are split evenly between control and the MTD. Thus choice of a fixed design for underlying models of this type is strongly indicated from these results.

If the underlying models are quadratic, the average SO design seems opposite to that for models which are non-quadratic. In this case it seems the best design would have most of the animals split evenly between the control group and the MTD dosed group and few animals at a very low middle dose ($d_1<.1$). However, the manifold describing the mean-squared-error of $\hat{s}$ over the range of designs is saddle-shaped for underlying quadratic dose-response models. One stirrup (or low-point of the saddle) is the SO design for quadratic models. The other stirrup is in the region of designs which are optimal for linear and linear-quadratic models. This implies that use of the designs which are optimal when the underlying model is linear or linear-quadratic will not cause a large increase in the mean-squared-error when the underlying model is quadratic.

Finally, for all underlying models, the designs seem to diverge as the MTD response and spontaneous background response increase. For the underlying quadratic and linear-quadratic models, this instability is due to changes in the design causing very little change in the value of the mean-squared-error. So, for these underlying models, most designs worked equally well and finding an optimal design is not as crucial. This is not the case when the underlying model is linear. For these models, as the spontaneous background increased, there was a definite trend toward an increased number of animals in the control group and an increase in

the value of d_1. Thus, prior knowledge concerning the probability of whether the underlying model is linear and what the background response will be would be beneficial in choosing an optimal design. It is assumed that no such knowledge is available in determining fixed optimal designs.

An optimal occurrence would be to find a design which never chooses the wrong model form. The SO designs for two cases in which the underlying model was linear had this property. When the underlying model is linear with .005 or .05 background response and 90% response at the MTD, using the SO designs given in Table 3.3 and fitting a linear-quadratic model, the resulting fitted model will be linear. Thus, this design has eliminated one of the two methods which add error to the estimation of the VSD. The resulting mean-squared-errors in Table 3.3 reflect this observation by being 5 to 7 orders of magnitude lower than what was observed for other designs. So, the effect of incorrect model choice on the estimated VSD is seen to be large and proper design can make significant improvements in estimation.

Consider now the differences in the AO (Table 2.4) and the SO (Table 3.3) designs. For both designs, the optimal middle doses (d_1) are similar when the underlying models are linear or linear-quadratic. For underlying quadratic models, the middle doses in the AO designs are consistently larger than those in the SO designs. In both designs, the number of animals at control increases as the background incidence increases. In the AO design, 50% to 70% of the animals are generally allocated to the middle dose with a majority of the remaining animals going into the control group. For linear and linear-quadratic underlying response, the SO designs seem to have this same trend. However,

when the underlying model is quadratic, the SO design puts most animals at the two endpoints, control and the MTD. In general, for underlying linear and linear-quadratic models, the AO designs seem close to the SO designs. When the underlying response is quadratic, the two seem quite different.

As was the case for the AO designs, the SO designs are model dependent and of little practical use to the laboratory scientist attempting to design a bioassay. What is required is a fixed design with doses expressed as proportions of the MTD, which works well for all underlying models. To find such a design, the relative efficiency of several designs to the SO (and other) designs will be used.

The relative efficiency of design ω' to some other design ω is defined as:

$$RE(\omega',\omega) = V_N(\omega;R,\varepsilon,P)/V_N(\omega';R,\varepsilon,P) \qquad (3.5.1)$$

where the mean-squared-error, V_N, is defined by (3.2.1). To find an optimal fixed design, the relative efficiencies of over one-hundred three-dose fixed designs were compared to the SO design. Under this routine, for a fixed design to be optimal, the value of the relative efficiency for that design compared to the SO design should be close to one for all underlying models.

Table 3.4 gives the relative efficiency of two fixed designs to the SO design for the 33 underlying models. The NCI design (the current NCI bioassay protocol) has a median relative efficiency of .380 and ranges in relative efficiency from .000 to .935. The remaining three-dose fixed design is the best three-dose fixed design found. It has a group given a dose of 30% the MTD using 90 animals and 30

TABLE 3.4: Relative Efficiency Of Three-Dose Fixed Designs To The SO Design

Back-ground	Model Type	MTD Response	doses animals	NCI 0.,.5,1. 50,50,50	ALT 0.,.3,1. 30,90,30
.005	LINEAR	.15		.297	.463
		.30		.227	1.000
		.50		.029	.780
		.90		.000	.000
	QUADRAT	.15		.184	.172
		.30		.138	.148
		.50		.136	.137
		.90		.267	.270
	LIN-QUA	.15		.330	.541
		.30		.352	.704
		.50		.143	.518
		.90		.654	.694
.050	LINEAR	.15		.348	.529
		.30		.588	.592
		.50		.380	.493
		.90		.000	.000
	QUADRAT	.15		.787	.690
		.30		.758	.714
		.50		.840	.833
		.90		.752	.839
	LIN-QUA	.15		.334	.413
		.30		.667	.709
		.50		.658	.909
		.90		.654	.910
.200	LINEAR	.30		.350	.403
		.50		.442	.535
		.90		.243	.345
	QUADRAT	.30		.806	.813
		.50		.935	.943
		.90		.917	.943
	LIN-QUA	.30		.392	.483
		.50		.599	.633
		.90		.855	.917

animals at both control and the MTD. This design will be referred to as the alternate or ALT design. The ALT design seems to perform more efficiently than the NCI design, especially when the underlying model of dose-response is linear or linear-quadratic. For this design, the median efficiency is .592 and the range in relative efficiency is from .000 to 1.000. From the NCI design to the ALT design, the drop in median relative efficiency is approximately .11 .

Concern over the very low relative efficiency in the underlying linear models with background incidence of .005 and .05 and response at the MTD of .90 is minimal. As was stated in the development of these models, MTD response generally ranges between 30% and 50% with most of all positive bioassays having MTD response in the range of 15% to 90%. Thus, if weights were assigned to these models, the 15% and 90% response models would receive little weight and the optimal design choice would correspond to SO designs in the MTD mid-response range.

From Table 3.4, it is seen that it is possible to improve the estimation of the VSD by changing the present design protocol. If the discussion were ended at this point, the recommended bioassay design would be the ALT design. However, there were assumptions made concerning the optimal design which must be investigated. These assumptions fixed the number of treated groups in the optimal design and forced the MTD to be one of the doses.

Recently, there has been much debate over the inclusion of a fourth dose in the cancer bioassay. Asymptotically, when using the two-stage model to do extrapolation, inclusion of a fourth dose would only lower the precision of the estimate of the VSD (Chernoff, 1953). However, for small bioassays, a fourth dose could help to eliminate the bimodality

of the estimate thus increasing its precision. There is also a bias expressed by the biologists, toxicologists, etc. toward the use of more doses. For these reasons, four-dose designs are studied.

Due to constraints on computer time and costs, it is impossible to attempt to find the four-dose optimal designs in the same manner in which the three-dose optimal designs (SO) were located. Instead, the concentration is on a set of four-dose designs which are indicated as being possibly optimal by the three-dose designs as well as designs being discussed in the toxicology literature. Tables 3.5, 3.6 and 3.7 portray a sampling of the better four-dose designs. In these tables, rather than give the relative efficiency of the four-dose designs to the SO designs, since the goal is to improve upon the ALT design, the mean-squared-error of the VSD for the ALT design will be used in the numerator of the relative efficiency. Therefore, for a four-dose design to improve upon the ALT design, the relative efficiencies expressed in Tables 3.5, 3.6 and 3.7 must be greater than 1.

In Table 3.5, the dose structure is one proposed by the Environmental Protection Agency (EPA) as a design which they feel will improve low-dose extrapolation. The animal allocations used were suggested by the EPA as well as the three dose optimal designs. Note that, in general, the first two animal allocations , labled AFD and BFD, do not perform as well as the ALT design. The CFD design, however, performs better than the ALT design for 10 of the 33 models. Seven of these models are for the high (90%) and low (15%) MTD response models. Thus, for the models of most frequent occurrence, the ALT design is superior. All three four-dose designs in Table 3.5 do significantly better than the ALT design for the two underlying linear models with 90% MTD

TABLE 3.5: Relatively Efficiency Of Four-Dose Fixed Designs To The ALT Design When Doses Of 0.,.2,.85 and 1. Are Employed

Back-ground	Model Type	MTD Response	animals	AFD 15,60,60,15	BFD 25,50,50,25	CFD 30,45,45,30
.005	LINEAR	.15		0.89	0.63	0.63
		.30		0.05	0.26	0.26
		.50		0.05	0.10	0.06
		.90		163486.40	20.65	114127.36
	QUADRAT	.15		1.11	1.28	1.27
		.30		0.88	0.96	0.90
		.50		1.10	1.06	1.11
		.90		1.02	0.98	1.00
	LIN-QUA	.15		0.75	0.48	0.49
		.30		0.70	0.65	0.74
		.50		0.38	0.40	0.36
		.90		1.02	0.98	1.01
.050	LINEAR	.15		0.79	0.70	0.69
		.30		0.51	0.63	0.66
		.50		0.34	0.41	0.52
		.90		2.44	39658.90	9945.32
	QUADRAT	.15		0.94	1.03	1.06
		.30		0.93	0.97	1.05
		.50		0.86	0.94	0.94
		.90		0.92	0.92	0.93
	LIN-QUA	.15		1.05	1.03	0.90
		.30		0.86	0.88	0.92
		.50		0.65	0.62	0.68
		.90		0.88	1.02	0.97
.200	LINEAR	.30		0.84	0.80	0.79
		.50		0.53	0.63	0.56
		.90		0.31	0.56	0.63
	QUADRAT	.30		1.03	1.01	1.11
		.50		0.92	0.95	1.01
		.90		0.88	0.85	0.88
	LIN-QUA	.30		0.71	0.68	0.69
		.50		0.68	0.66	0.75
		.90		0.96	1.09	1.09

TABLE 3.6: Relatively Efficiency Of Four-Dose Fixed Designs To The ALT Design When Doses Of 0.,.33,.67 and 1. Are Employed

Back-ground	Model Type	MTD Response	animals	DFD 15,60,60,15	EFD 25,50,50,25	FFD 30,45,45,30
.005	LINEAR	.15		0.88	0.91	0.88
		.30		0.30	0.28	0.26
		.50		0.09	0.10	0.15
		.90		1979.45	16295.56	7.69
	QUADRAT	.15		1.25	1.08	1.16
		.30		1.01	1.04	1.05
		.50		1.09	0.99	1.03
		.90		0.97	0.98	1.02
	LIN-QUA	.15		0.75	0.79	0.71
		.30		0.59	0.70	0.70
		.50		0.40	0.41	0.64
		.90		1.04	1.05	0.97
.050	LINEAR	.15		1.19	0.89	0.95
		.30		0.82	1.02	0.97
		.50		0.59	1.00	0.98
		.90		11.11	5.00	6.66
	QUADRAT	.15		0.94	1.05	1.08
		.30		0.96	1.08	1.05
		.50		0.88	0.90	1.00
		.90		0.88	0.96	0.93
	LIN-QUA	.15		1.33	1.03	0.94
		.30		1.06	0.94	1.03
		.50		0.67	0.82	0.89
		.90		0.91	0.88	0.81
.200	LINEAR	.30		1.41	1.28	1.11
		.50		0.91	0.97	1.09
		.90		0.80	1.43	1.08
	QUADRAT	.30		0.94	1.00	0.96
		.50		0.93	0.96	1.03
		.90		0.87	0.86	0.89
	LIN-QUA	.30		1.09	1.02	1.00
		.50		0.95	0.93	0.81
		.90		1.02	0.99	1.00

TABLE 3.7: Relatively Efficiency Of Four-Dose Fixed Designs To The ALT Design When Doses Of 0.,.1,.33 and 1. Are Employed

Back-ground	Model Type	MTD Response	animals	GFD 15,60,60,15	HFD 25,50,50,25	IFD 30,45,45,30
.005	LINEAR	.15		0.65	0.63	0.63
		.30		0.20	0.16	0.20
		.50		0.03	0.10	0.08
		.90		42.02	100633.99	24198.43
	QUADRAT	.15		1.23	1.23	1.22
		.30		1.27	1.18	1.12
		.50		1.19	1.11	1.06
		.90		1.19	0.95	1.09
	LIN-QUA	.15		0.53	0.50	0.51
		.30		0.41	0.53	0.51
		.50		0.28	0.48	0.46
		.90		0.99	1.05	1.03
.050	LINEAR	.15		0.90	0.79	0.72
		.30		0.86	0.87	0.72
		.50		0.39	0.60	0.68
		.90		1.85	3.33	12.50
	QUADRAT	.15		1.06	1.05	1.06
		.30		1.10	1.01	1.09
		.50		0.88	0.88	0.95
		.90		0.91	0.91	0.95
	LIN-QUA	.15		1.23	0.96	0.93
		.30		0.89	0.92	0.85
		.50		0.63	0.69	0.67
		.90		0.94	0.88	1.01
.200	LINEAR	.30		1.25	1.09	0.83
		.50		0.98	0.85	0.70
		.90		0.44	0.72	2.27
	QUADRAT	.30		0.91	0.97	1.01
		.50		0.95	0.94	0.95
		.90		0.93	1.01	0.92
	LIN-QUA	.30		1.19	0.90	0.72
		.50		1.16	0.93	0.55
		.90		0.88	0.93	1.03

response and .005 or .05 spontaneous background incidence. Thus, these four-dose designs more closely approximate the SO designs when the design yields "perfect" response modelling.

In Table 3.6 the dose structure mimics that used in the NCI bioassay protocol; the doses are uniformly spaced between 0 and 1. The designs in this table tend to do slightly better than those of Table 3.5. However the DFD and EFD designs still do not perform consistently better than the ALT design. The FFD design has a smaller mean-squared-error than the ALT design for 13 models and ties with it for 3 models. So, for roughly half of the models, the ALT design can do no better than the FFD design. Other considerations will be needed to distinguish between these designs.

The remaining four-dose designs, shown in Table 3.7, have two moderately small middle doses. The object here was to determine what would be the result of splitting the middle dose of the ALT design which was at .3. The resulting designs, GFD, HFD and IFD, have increased efficiency for the quadratic models and linear models with 90% MTD response. The loss in efficiency for the remaining models is typically about a two-fold increase in the mean-squared-error. These designs do not seem to improve the extrapolation when compared to the ALT design.

The final assumption to be tested concerns the use of the MTD as the upper dose in the optimal design. It is possible that there exists some design with d_2<MTD which produces a smaller mean-squared-error for $\hat{s}$. In Table 3.8, four of the many possible non-MTD designs are considered. The table entries are the relative efficiency of the non-MTD designs to the ALT design for each underlying model.

If the upper dose in the ALT design is slowly dropped below the

TABLE 3.8: Relative Efficiency Of Three-Dose Non-MTD Fixed Designs To The ALT Design

Back-ground	Model Type	MTD Response	doses animals	0.,.3,.9 30,90,30	0.,.3,.8 30,90,30	0.,.25,.5 30,90,30	0.,.25,.5 50,50,50
.005	LINEAR	.15		1.18	1.61	1.05	0.40
		.30		0.64	1.47	0.40	0.19
		.50		1079.88	1.12	4.76	0.15
		.90		1.39	3.23	1260.27	1246.89
	QUADRAT	.15		0.79	0.75	0.56	0.55
		.30		0.81	0.66	0.43	0.39
		.50		0.79	0.62	0.45	0.43
		.90		0.67	0.52	0.46	0.47
	LIN-QUA	.15		1.00	1.16	0.72	0.30
		.30		1.56	1.39	1.02	0.64
		.50		1.67	2.78	1.64	2.78
		.90		1.52	2.94	16.67	5.26
.050	LINEAR	.15		1.11	0.99	1.45	0.88
		.30		1.03	1.03	1.56	0.70
		.50		1.72	2.86	1.11	1.11
		.90		2.00	4.00	103.88	540.10
	QUADRAT	.15		0.90	0.88	0.74	0.76
		.30		0.95	0.88	0.65	0.65
		.50		0.85	0.67	0.50	0.47
		.90		0.66	0.53	0.43	0.43
	LIN-QUA	.15		1.15	1.14	1.39	0.92
		.30		1.10	1.04	1.41	0.76
		.50		0.97	1.08	0.88	0.76
		.90		1.72	3.13	5.03	3.03
.200	LINEAR	.30		1.28	1.47	2.63	1.49
		.50		0.93	1.08	1.43	0.88
		.90		0.98	4.35	1.09	2.38
	QUADRAT	.30		0.89	0.87	0.72	0.75
		.50		0.89	0.76	0.61	0.64
		.90		0.74	0.60	0.43	0.40
	LIN-QUA	.30		0.96	1.19	2.70	1.45
		.50		0.91	1.12	2.02	1.08
		.90		1.20	1.08	1.06	0.89

MTD, the pattern of change in the mean-squared-error can be observed. From Table 3.8 it is seen that by dropping the upper dose to 90% of the MTD, for the linear and linear-quadratic models, the efficiency of the estimation improves. By going a little farther and decreasing this dose to 80% of the MTD, the estimation of the VSD is improved for all but one of the underlying linear and linear-quadratic models. For quadratic models, as the upper dose is lowered, the relative efficiency decreases indicating that the precision of the estimation of the VSD is decreasing. When the upper dose decreases to 50% of the MTD, the efficiency increases for some models and decreases for others indicating that lower movement will not aid precision. For the quadratic models, the deterioration of efficiency continues as the high dose is lowered to 50% of the MTD. This loss for the quadratic models and gain for the linear and linear-quadratic models is directly due to the probability of the linear-quadratic model fitting a linear model. As the high dose decreases, this probability increases thus making the estimation of the VSD more accurate if the underlying model is linear or linear-quadratic and less accurate if the underlying model is quadratic.

Under the assumption of a three-dose design with the upper dose fixed at the MTD, being optimal, it is seen that use of the ALT design will most improve the low-dose extrapolation when compared to any other design. If four-dose designs are considered, the FFD design of Table 3.6 is a viable alternative to the ALT design. Finally, if no other mathematical or biological considerations are included, decreasing the upper dose in the ALT design to 80% of the MTD can significantly improve estimation if the underlying dose-response model is linear or linear-quadratic at the expense of increasing the mean-squared-error if

the underlying dose-response model is quadratic.

The improvement in extrapolation from use of the ALT design can be explained. The group at 30% MTD does better at distinguishing between linear and quadratic response than a dose at 50% the MTD. By putting more animals at this point, the difference between the models is better defined. Four-dose designs improve upon this by averaging out the variability of this one low dose and thus, doing a better job of distinguishing between model forms. Finally, the non-MTD designs improve upon this by lowering the upper dose to a point which improves the detection for linear models. However, if this upper dose is lowered too much, all response looks linear and a quadratic model is seldom fit.

CHAPTER IV

POWER CONSIDERATIONS IN BIOASSAY DESIGN

4.1 Introduction

The discussion in Chapter I deals with the cancer bioassay and the two distinct functions generally associated with its use; qualitative and quantitative risk assessment. Up to this point, the discussion has centered upon the quantitative aspects of risk assessment and how changes in the bioassay design will affect the accuracy of low-dose extrapolation under the multi-stage model.

In Chapter II, the discussion centered on the asymptotic theory of extrapolation and designs which optimize the low-dose extrapolation for bioassays utilizing an infinite number of animals. In this situation, qualitative risk assessment plays a very minor role in the results. It would be very difficult, if not impossible, to combine the two for asymptotically large bioassays. The AO designs developed in Chapter II are assumed to be applied to experiments on chemicals which are known, a priori, to be carcinogenic.

In Chapter III, a set of theory is developed concerning the action of low-dose extrapolation for small bioassays. However, in this case, the extrapolations were censored and only performed when a significant dose-response trend was detected. Again the extrapolation process is assumed to be applied only for experiments on chemicals which have

shown a positive tumor induction. Several designs seemed to improve upon the current bioassay design for low-dose extrapolation and it was difficult to choose between the better of these designs.

This approach as the final word in developing a usable cancer bioassay is unacceptable. It is important to remember that, if because of faulty design, it is not possible to detect a positive dose-response when one exists, then that design is of very little use as a practical protocol. If the designs of Chapter III have this characteristic, their optimality properties are wasted since it is highly unlikely extrapolation will ever be used to estimate the VSD. The use of two bioassays, one for qualitative risk assessment and one for quantitative risk assessment, is also unacceptable.

The property being discussed is the power of the test used to detect increased response. For a two-dose design (one dose and the control group) the most commonly used test is the Fisher Exact Test. In bioassays with three or more doses, the appropriate test is the Armitage linear trend test. These were discussed in more detail in Chapter I.

A second criterion which could be used to modify the bioassay design would be to maximize the power of the test for increased tumor incidence. However, the answer for this design is simple and intuitive. For both the Fisher and Armitage tests, the most powerful design against any alternative (or underlying design) has two groups, control and MTD, with 50% of the animals in each group. This statement has been shown true by the simulation study for small bioassays. For large bioassays, this design is derived by use of the asymptotic power of the Armitage test (Chapman and Nam, 1968).

Instead of simply finding an optimal design by minimizing the mean-squared-error or maximizing the power of the test for linear trend, the goal of this chapter is to combine the two criteria to develop a design which is both practical and useful. In the remaining sections of this chapter, several alternative occurrences which will affect the combination of qualitative and quantitative risk assessment in developing an optimal bioassay design are discussed.

4.2 Small Bioassay Power For Linear Trend Tests

The simplest method of combining power for detection and mean-squared-error for estimation is to simply study the power for the better designs of Chapter III. A rather loose criterion in this case would be to choose that design which yields the highest power for the better designs. In the simulation study, the experiments were censored by requiring a significant increase in incidence. This was defined as passing the Armitage test at the .05 significance level. By simply counting the number of times increased incidence was significant and dividing by the number of simulated experiments, a measure of the power of the Armitage test at the .05 significance level is obtained. The alternative hypothesis under such a procedure will be the underlying model from which the simulated experiments are generated.

Table 4.1 gives the power as described above for the SO design, the NCI design and the ALT design. In general, the power decreases as the response at the MTD decreases. This is due to the fact that, the lower the MTD response is, the less steep the increased incidence and the less likely the chances of detecting a significant increase. Similarly, the power decreases as the background incidence increases.

TABLE 4.1: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For Three-Dose Designs

Back-ground	Model Type	MTD Response	SO	NCI	ALT
.005	LINEAR	.15	.52	.90	.72
		.30	.97	.99	.96
		.50	.99	1.00	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.98	.94	.90
		.30	1.00	1.00	.99
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.59	.93	.78
		.30	.79	1.00	.97
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
.050	LINEAR	.15	.26	.52	.42
		.30	.60	.96	.86
		.50	.98	1.00	.99
		.90	1.00	1.00	1.00
	QUADRAT	.15	.68	.55	.52
		.30	1.00	.98	.94
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.33	.54	.49
		.30	.64	.96	.91
		.50	.91	1.00	1.00
		.90	1.00	1.00	1.00
.200	LINEAR	.30	.19	.31	.28
		.50	.66	.94	.86
		.90	1.00	1.00	1.00
	QUADRAT	.30	.41	.31	.27
		.50	.98	.95	.91
		.90	1.00	1.00	1.00
	LIN-QUA	.30	.16	.30	.28
		.50	.68	.95	.89
		.90	1.00	1.00	1.00

For the SO design, the power for detection is inadequate for linear and linear-quadratic models. Recall that these designs had most of the animals at about 30% of the MTD. This dose is too low to show a significant increase in tumor incidence when the MTD response is small. Thus, for this case, the ideal situation of the optimal design for extrapolation being the optimal design for testing is not obtained.

For the underlying quadratic models, the situation is quite different. In this case, the power for detection of increased incidence is quite high. In fact, subject to the constraints imposed in the grid search, these designs yield the highest power possible. Thus the optimal design for extrapolation is also the optimal design for testing.

Consider the power for the NCI bioassay protocol. If the design is to be modified, it is desirable to increase the power for detection (or at least maintain it) while simultaneously decreasing the mean-squared-error of the estimate of the VSD over that for the NCI design. In general, the NCI design is quite powerful. For all expected response in which the difference between the control group response and the MTD group response is greater than 15%, the power for detection is greater than 90%. In the case of quadratic models, the NCI design is almost as powerful as the most powerful design (SO design).

The three-dose design which was considered the best for quantitative risk assessment is less powerful than the NCI design . For the underlying quadratic models, the loss is not as bad as when the underlying model is linear or linear-quadratic. Previously, it was seen that the largest gains in decreasing the mean-squared-error of the VSD occurred for underlying linear and linear-quadratic models. Thus, where the ALT design is optimal for extrapolation, it has very low power for

testing. In general, using power as the criterion for optimality, the NCI design would be chosen over the ALT design.

Now, consider the nine four-dose designs examined for quantitative risk assessment in Chapter III. In Table 4.2 the power of the Armitage test at the .05 significance level for the four-dose designs using doses at control, 20% MTD, 85% MTD and the MTD are presented. Regardless of how the animals are allocated, these designs are as powerful, or more powerful, than the NCI design. Although the MSE($\hat{s}$) is generally smaller for the ALT design than it is for the CFD design, the CFD design did perform better than the NCI design for estimation of the VSD. Thus including power has clouded the choice between the CFD and the ALT designs.

Table 4.3 gives the power of the Armitage test for the four-dose designs using doses at control, 33% MTD, 67% MTD and the MTD. In comparison to the results of Table 4.2, a decrease in power is detected. Thus, by increasing d_1 and decreasing d_2 the power for increased risk detection is lowered. The FFD design performed as well as the ALT design for extrapolation. From Table 4.4, it is seen that they also have the same power for detection, thus differentiation between the ALT and FFD designs is again unattainable.

For the final three four-dose designs of Table 4.5, the power increases as the two middle doses are lowered to 10% MTD and 33% MTD. Thus, even though these designs did not perform as well as the other designs for extrapolation, the power for detection is high and they cannot be disregarded for being inferior.

Finally, consider the power of the non-MTD designs of Chapter III for the 33 underlying models shown in Table 4.5. Here, it is seen that as the highest dose decreases, the power decreases to a point where, for

TABLE 4.2: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For Four-Dose Designs With Doses 0., .20, .85 and 1.

Back-ground	Model Type	MTD Response	AFD	BFD	CFD
.005	LINEAR	.15	.82	.85	.88
		.30	1.00	1.00	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.93	.94	.97
		.30	1.00	1.00	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.84	.88	.92
		.30	.99	1.00	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
.050	LINEAR	.15	.42	.51	.50
		.30	.91	.94	.95
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.49	.53	.54
		.30	.96	.98	.98
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.48	.54	.54
		.30	.92	.96	.96
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
.200	LINEAR	.30	.26	.29	.31
		.50	.87	.92	.93
		.90	1.00	1.00	1.00
	QUADRAT	.30	.29	.31	.33
		.50	.93	.96	.96
		.90	1.00	1.00	1.00
	LIN-QUA	.30	.27	.31	.31
		.50	.88	.93	.95
		.90	1.00	1.00	1.00

TABLE 4.3: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For Four-Dose Designs With Doses 0., .33, .67 and 1.

Back-ground	Model Type	MTD Response	DFD	EFD	FFD
.005	LINEAR	.15	.55	.70	.75
		.30	.88	.96	.98
		.50	.99	1.00	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.67	.82	.86
		.30	.96	.99	.99
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.57	.73	.78
		.30	.87	.97	.99
		.50	.99	1.00	1.00
		.90	1.00	1.00	1.00
.050	LINEAR	.15	.28	.37	.40
		.30	.71	.83	.87
		.50	.95	.99	.99
		.90	1.00	1.00	1.00
	QUADRAT	.15	.31	.42	.45
		.30	.81	.90	.93
		.50	.99	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.32	.42	.43
		.30	.75	.97	1.00
		.50	.98	1.00	1.00
		.90	1.00	1.00	1.00
.200	LINEAR	.30	.20	.23	.25
		.50	.66	.80	.84
		.90	1.00	1.00	1.00
	QUADRAT	.30	.20	.23	.26
		.50	.72	.84	.88
		.90	1.00	1.00	1.00
	LIN-QUA	.30	.18	.23	.26
		.50	.66	.83	.85
		.90	1.00	1.00	1.00

TABLE 4.4: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For Four-Dose Designs With Doses 0., .10, .33 and 1.

Back-ground	Model Type	MTD Response	GFD	HFD	IFD
.005	LINEAR	.15	.67	.80	.87
		.30	.92	1.00	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.82	.92	.95
		.30	.98	1.00	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.71	.82	.89
		.30	.92	.98	1.00
		.50	1.00	1.00	1.00
		.90	1.00	1.00	1.00
.050	LINEAR	.15	.35	.43	.52
		.30	.78	.86	.94
		.50	.97	.99	1.00
		.90	1.00	1.00	1.00
	QUADRAT	.15	.40	.52	.56
		.30	.84	.92	.98
		.50	.99	1.00	1.00
		.90	1.00	1.00	1.00
	LIN-QUA	.15	.40	.46	.59
		.30	.80	.91	.96
		.50	.98	1.00	1.00
		.90	1.00	1.00	1.00
.200	LINEAR	.30	.20	.25	.29
		.50	.69	.83	.93
		.90	1.00	1.00	1.00
	QUADRAT	.30	.24	.27	.32
		.50	.75	.90	.93
		.90	1.00	1.00	1.00
	LIN-QUA	.30	.22	.27	.33
		.50	.73	.83	.93
		.90	1.00	1.00	1.00

TABLE 4.5: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For Non-MTD Designs

Back-ground	Model Type	MTD Response	doses animals	0.,.3,.9 30,90,30	0.,.3,.8 30,90,30	0.,.25,.5 30,90,30	0.,.25,.5 50,50,50
.005	LINEAR	.15		.67	.65	.36	.61
		.30		.94	.93	.74	.95
		.50		1.00	1.00	.96	1.00
		.90		1.00	1.00	1.00	1.00
	QUADRAT	.15		.82	.64	.25	.41
		.30		.98	.95	.52	.77
		.50		1.00	1.00	.82	.95
		.90		1.00	1.00	1.00	1.00
	LIN-QUA	.15		.72	.66	.37	.61
		.30		.95	.92	.68	.93
		.50		1.00	1.00	.93	.99
		.90		1.00	1.00	1.00	1.00
.050	LINEAR	.15		.36	.32	.18	.25
		.30		.81	.74	.47	.70
		.50		.99	.97	.85	.97
		.90		1.00	1.00	.96	1.00
	QUADRAT	.15		.39	.31	.12	.15
		.30		.89	.74	.27	.37
		.50		.99	.98	.61	.74
		.90		1.00	1.00	.99	1.00
	LIN-QUA	.15		.39	.31	.12	.16
		.30		.82	.74	.37	.52
		.50		.98	.97	.74	.91
		.90		1.00	1.00	1.00	1.00
.200	LINEAR	.30		.22	.19	.12	.15
		.50		.73	.68	.43	.58
		.90		1.00	1.00	1.00	1.00
	QUADRAT	.30		.24	.18	.08	.10
		.50		.81	.62	.20	.25
		.90		1.00	1.00	.86	.96
	LIN-QUA	.30		.21	.20	.11	.15
		.50		.76	.68	.33	.43
		.90		1.00	1.00	.89	.98

certain underlying models, the power is less than 10%. The non-MTD design which seemed optimal for extrapolation (d_2=.80) has much lower power than the ALT design and the best four-dose designs. For some underlying models, the loss in power compared to the ALT design is greater than 30%. The decrease in the mean-squared-error for $\hat{s}$ in these designs is offset by the loss in power. Therefore, non-MTD designs do not seem to improve the bioassay and will be dropped.

4.3 Loss Of The Maximum Tolerated Dose

As defined in Chapter I and used in practice, the MTD is an estimate of the highest dose which is not acutely toxic to the test animal. Because the MTD is an estimate, it is possible to overestimate the MTD and either kill all the animals at the highest administered dose or to kill (or otherwise affect) enough animals to make the information attained at this dose of little use in the analysis. Also, besides death, there are other toxic effects not related to carcinogenesis which can affect response at the high dose. In these cases, it is possible to observe increased incidence at the low-dose group and no increased incidence at the high-dose group. The standard procedure for such an occurrence would be to determine the agent non-carcinogenic. These problems are referred to as loss of the MTD. In terms of the NCI Carcinogenesis Bioassay Program, loss of the MTD group will occur approximately 15% of the time.

Loss of the MTD is of major concern to biologists and toxicologists when they request more doses be used in the bioassay design. Obviously, loss of a dosed group will significantly affect the risk assessment of any one chemical. In terms of the power of the test for carcinogenesis,

depending on the number of animals in the loss group, decreases in power can be expected. Table 4.6 gives the power of the test for increased incidence for the better designs used in extrapolation when the MTD dosed group is lost.

For the NCI and ALT designs, the appropriate test becomes the Fisher test since, only a control group and a treated group remain. For the NCI design, if the underlying model had low expected response at the MTD, the loss in power is about 40% to 90%. For higher MTD response , there is very little loss in power. The ALT design has drastic power loss of greater than 60% for underlying models with expected MTD response as high as .90. In some cases, the power is effectively zero. Thus, when loss of the MTD group occurs, this design would most likely be unable to detect an increased tumor incidence.

For four-dose designs, when the MTD is lost, two dosed groups and a control group remain, so the appropriate test is the Armitage linear trend test. In Table 4.6, the power of the Armitage linear trend test at the .05 significance level is shown for the two best four-dose designs; the CFD design and the FFD design. For both designs, the power remains very high. For linear and linear-quadratic models, the two designs seem very close, the CFD design has a slightly lower power. For the underlying quadratic models, the CFD design does much better in terms of power, than the FFD design.

So, it is observed that, if the MTD group is lost, the four-dose designs can protect the validity of the experiment for testing. The ALT design, although optimal for extrapolation, would most likely fail to detect increased incidence if the MTD group were lost. Similar comments apply for the NCI design. Chapter V will summarize these observations

TABLE 4.6: Small Bioassay Power Of The Armitage Linear Trend Test At The .05 Significance Level For loss Of The MTD

Back-ground	Model Type	MTD Response	NCI*	ALT*	FFD	CFD
.005	LINEAR	.15	.31	.01	.65	.78
		.30	.89	.40	.96	,98
		.50	.99	.94	.99	1.00
		.90	1.00	1.00	1.00	1.00
	QUADRAT	.15	.05	.00	.52	.82
		.30	.39	.00	.90	.99
		.50	.88	.06	1.00	1.00
		.90	1.00	.94	1.00	1.00
	LIN-QUA	.15	.26	.01	.64	.80
		.30	.83	.26	.93	.97
		.50	.99	.78	.99	1.00
		.90	1.00	.96	1.00	1.00
.050	LINEAR	.15	.14	.05	.28	.37
		.30	.58	.22	.76	.86
		.50	.96	.65	.98	.99
		.90	1.00	1.00	1.00	1.00
	QUADRAT	.15	.06	.02	.19	.37
		.30	.22	.04	.56	.86
		.50	.62	.10	.91	1.00
		.90	1.00	.68	1.00	1.00
	LIN-QUA	.15	.06	.02	.24	.34
		.30	.39	.10	.66	.88
		.50	.85	.34	.96	1.00
		.90	1.00	.87	1.00	1.00
.200	LINEAR	.30	.10	.07	.16	.24
		.50	.50	.22	.65	.81
		.90	1.00	.97	1.00	1.00
	QUADRAT	.30	.07	.06	.12	.20
		.50	.19	.07	.46	.78
		.90	.95	.32	1.00	1.00
	LIN-QUA	.30	.10	.07	.16	.23
		.50	.34	.13	.59	.80
		.90	.97	.44	1.00	1.00

* Power based on the Fisher Exact Test at .05 significance

and discuss other possible methods of including power in determining an optimal bioassay design.

CHAPTER V

SUMMARY AND COMMENTS FOR FUTURE RESEARCH

5.1 Summary

Chapter II dealt with the asymptotic theory concerning the estimation of the VSD when the underlying dose-response follows a multi-stage model. The distribution of the estimate of the VSD was derived for the two-stage model using large sample theory. It was shown that, due to the constrained nature of the parameters in the multi-stage model, these distributions could be non-normal and the estimates biased. The formula for the mean-squared-error of the VSD was derived.

Using historical dose-response, a set of linear, quadratic and linear-quadratic models was developed. The mean-squared-error of the VSD was minimized under the assumption of each of these models yielding a set of asymptotically optimal designs. These designs were seen to be independent of the added risk, ε, and certain trends in design were detected concerning spontaneous background incidence and response at the MTD. It was noted that the asymptotically optimal (AO) designs are model dependent and the practical need is to develop fixed designs which are based only on knowledge of the MTD.

Asymptotic relative efficiency (ARE) was used to compare several fixed designs to the AO designs. One design had a larger relative efficiency than the NCI design for all 33 models being investigated. This design is clearly an improvement over the NCI design for large

bioassays. One other design had an asymptotic relative efficiency larger than .5 in all 33 models making this design consistency close to the AO design. Both of these designs could be considered optimal fixed bioassay designs.

Chapter III dealt with optimal designs in bioassays using small numbers of animals. For small bioassays, Monte Carlo simulation techniques were utilized to derive the distribution of the estimate of the VSD and to minimize the mean-squared-error of the VSD. It was shown that by fitting a linear or a quadratic model, the estimation of the VSD could be orders of magnitude off if the underlying model was of a different form. This was due to the fact that the slope of the dose-response curve at a dose of zero is zero for the quadratic model and non-zero for the linear model.

The remainder of Chapter III dealt with fitting the linear-quadratic model. It was shown that, regardless of the underlying model, the distribution of the estimated VSD was bimodal for small bioassays. For some underlying models, this bimodality remained for bioassays with as many as 100,000 animals.

The distribution of the VSD derived using large sample theory was compared to what was observed from the simulation study. For underlying linear and linear-quadratic models, the large sample distribution did not correctly approximate the right-hand tail of the distribution of the VSD for small bioassays. Similarly, the left-hand tail approximation was poor when the underlying dose-response model was quadratic. In general, it was shown that the asymptotic theory of Chapter II could not be used to approximate the distribution of the VSD for small bioassays.

Several methods of approximating confidence bounds were compared to confidence bounds observed in the simulation study. In general, none

of the methods worked well as upper confidence bounds. The method proposed by Crump yielded lower confidence bounds which were generally conservative yet close enough to the observed lower bounds to be acceptable. In all cases, the asymptotic theory of Chapter II yielded unacceptable confidence bounds.

Employing simulation methodology and a directed grid search, optimal designs for bioassays utilizing 150 animals were derived for each of the 33 models created in Chapter II. These designs were seen to be robust to the added acceptable risk and moderately insensitive to response at the MTD and spontaneous incidence. The optimal designs for linear and linear-quadratic underlying models were all similar and those for quadratic models were the same. However, due to the saddle shape of the manifold describing the mean-squared-error of the estimated VSD, the optimal design structure for linear and linear-quadratic models was near optimal for quadratic models. Finally, it was shown that it is sometimes possible to modify the design such that, when a linear-quadratic model is fit to observed response, it will fit the correct model form almost 100% of the time.

Again, these optimal designs were model dependent and we wished to locate fixed designs which were near optimal for all underlying models. The ALT design of Table 3.4 was the best fixed three-dose design utilizing the MTD as one of the doses. Several four-dose designs were compared to the ALT design in terms of the mean-squared-error of the VSD. Two four-dose designs seemed to perform as well as the ALT design for low-dose extrapolation. These were the CFD design and the FFD design. The final set of fixed designs compared to the ALT design were designs which did not utilize the MTD as one of the doses. It was seen that by dropping

the highest dose in the ALT design from the MTD to 80% of the MTD, the extrapolation could be improved when the underlying model was linear or linear-quadratic at the expense of increasing the variability of the estimate of the VSD when the underlying model was quadratic. Lowering the high-dose to below 80% of the MTD did not seem to improve the extrapolation.

In Chapter IV, the power of the test for linear trend was considered in choosing an optimal bioassay design. It was shown that by not using the MTD as the highest dose group, the power dropped dramatically. The NCI design had very good power and was more powerful than the ALT design. The two four-dose designs which performed well for extrapolation, the CFD and FFD designs, were as powerful as the NCI design with the CFD design being slightly more powerful. When toxicity in the high-dose group was considered, it became clear that the CFD design was the most powerful design considered, followed by the FFD design. The ALT design and NCI design displayed dramatic power loss when the MTD dosed-group was unusable.

Either of the two four-dose designs, CFD and FFD, would improve the bioassay for risk assessment. The CFD design is slightly more powerful than the FFD design for qualitative risk assessment. The FFD design is slightly more efficient than the CFD design for quantitative risk assessment. The CFD design has doses at 0 (control), 20% of the MTD, 85% of the MTD and the MTD. The FFD design has doses at 0 (control), 33% of the MTD, 67% of the MTD and the MTD. Both designs use 30 animals at control and the MTD and 45 animals at each of the mid-range doses. Any design falling between these two designs (in terms of the middle doses) would also be acceptable.

5.2 Future Research

Several additional aspects of the problem of optimal bioassay design could be studied. We have studied the problem of optimal bioassay design when it is assumed that the mechanism of carcinogenic response could be described by a multi-stage theory. The same methodology could be employed to find optimal designs using other dose-response models such as the Weibull, probit, logit and multi-hit models as well as higher order multi-stage models. The asymptotic theory could also be expanded to include the general k-stage model.

In addition to the properties considered here pertaining to the optimal bioassay design, it would also be of interest to study the false-positive rate of the test for increased tumor incidence using the optimal design. Simply put, suppose an agent is non-carcinogenic and a bioassay is performed to test the carcinogenicity of this agent. An important question is "what is the probability of determining this agent to be a positive carcinogen given it is not a carcinogen?" The design of the bioassay will play a key role in the answer to this question.

A more difficult problem is optimal bioassay design when parametric time-to-tumor models such as the proportional hazards model and the general product model are used. The same methodology used for the multi-stage model could be employed under these models. Additional variables such as intermediate sacrifice times could be incorporated into such a design problem. Several questions remain unanswered for time-to-tumor models. There is much debate concerning the definition of the VSD when a time-to-tumor model is employed in the extrapolation. The most important question concerns the efficacy of using time-to-tumor models for quantitative risk assessment. Many authors feel that the

incorporation of time in the model does not aid in the extrapolation and may possibly make the process more variable.

The choice of an optimal design in this research was based upon the use of a symmetric loss function. It was assumed that underestimation of the VSD was as serious as overestimation. It has been proposed that a non-symmetric loss function be utilized. Such a function would be generally unacceptable with either the regulatory agencies or industry objecting to the weighting scheme. However, if such a loss function were to be accepted, it could be applied in the same manner in which we applied the mean-squared-error to locate optimal designs.

Finally, all of the results presented here have been based upon extrapolation to the VSD. Several alternative measures of the characterization of risk have been proposed. Methods of optimizing the bioassay design to improve upon the estimation of some other characterization of risk could be employed.

BIBLIOGRAPHY

Albert, R. and Altshuler, B., Considerations Relating To The Formulation Of Limits For Unavoidable Population Exposures To Environmental Carcinogens, Radionuclide Carcinogenesis, AEC Symposium Series, Conf.-720505, Springfield, Virginia (1969) pp. 233-253

Armitage, P., Statistical Methods In Medical Research, New York, Wiley (1971)

Armitage, P. and Doll, R., The Age Distribution Of Cancer And A Multi-Stage Theory Of Carcinogenesis, British Journ. of Cancer 8 (1954) pp. 1-12

Bard, Y., Nonlinear Parameter Estimation, New York, Academy (1974)

Campbell, T., Chemical Carcinogens And Human Risk Assessment, FASEB Proceedings 39 (1980) pp. 2467-2484

Chand, N. and Hoel, D., A Comparison Of Models For Determining Safe Levels Of Environmental Agents, Reliability and Biometry, SIAM Phil. (1974) pp. 681-700

Chapman, D. and Nam, J., Asymptotic Power of Chi-Square Tests For Linear Trends In Proportions, Biometrics 22 (1968) pp. 315-327

Chernoff, H., Locally Optimal Designs For Estimating Parameters, Ann. Math. Stat. 24 (1953) pp. 586-602

Cornfield, J., Carlborg, F. and Van Ryzin, J., Setting On The Basis Of Mathematical Treatment Of Dose-Response Data Extrapolated To Low Doses, Proc. First Intl. Tox. Congress (1978)

Crump, K., An Improved Procedure For Low-Dose Carcinogenic Risk Assessment From Animal Data, (To Appear In) J. of Env. Path. Tox. (1980)

Crump, K., Hoel, D., Langley, C. and Peto, R., Fundamental Carcinogenic Processes And Their Implications For Low Dose Risk Assessment, Cancer Research 36 (1976) pp. 2973-2979

Falk, H., Biologic Evidence For The Existence Of Thresholds In Chemical Carcinogens, Env. Health Perspectives 22 (1978) pp. 167-170

Gart, J., Chu, K. and Tarone, R., Statistical Issues In Interpretation Of Chronic Bioassay Tests For Carcinogenicity, J. Nat. Cancer Inst. 62 (1979) pp. 957-974

Gaylor, D. And Hoel, D., Statistical Analysis Of Carcinogenesis Data From Chronic Animal Studies, in Sontag, J. (ed.), Carcinogens In Industry And Environment, Dekker, New York (1981) pp.97-111

Gue, R. and Thomas, M., Mathematical Methods In Operations Research, London, Macmillan (1968)

Guess, H. and Crump, K., Best-Estimate Low-Dose Extrapolation Of Carcinogenicity Data, Env. Health Persp. 22 (1978) pp. 149-152

Guess, H. and Crump, K., Low-Dose-Rate Extrapolation Of Data From Animal Carcinogenicity Experiments - Analysis Of A New Statistical Technique, Biosciences 32 (1976) pp. 15-36

Guess, H., Crump, K. and Peto, R., Uncertainty Estimates For Low-Dose-Rate Extrapolations Of Animal Carcinogenicity Data, Cancer Research 37 (1977) pp. 3475-3483

Guess, H. and Hoel, D., The Effect Of Dose On Cancer Latency Period, J. of Envr. Path. and Tox. 1 (1977) pp. 279-286

Hartley, H. and Sielken, R., Estimation Of "Safe Doses" In Carcinogenic Experiments, Biometrics 31 (1977) pp. 1-30

Haseman, J., Hoel, D. and Jennrich, R., Some Practical Problems Arising From Use Of The Gamma Multi-Hit Model For Risk Estimation, (submitted to) J. of Tox. and Envr. Health (1980)

Hoel, D., Incorporation Of Background In Dose-Response Models, Fed. Proc. 39 (1980) pp. 73-75

Hoel, D., Gaylor, D., Kirchstein, R. Umberto, S. and Schneiderman, M., Estimation Of Risks Of Irreversible, Delayed Toxicity, J. of Tox. and Envr. Helath 1 (1975) pp. 133-151

Hoel, P. and Jennrich, R., Optimal Designs For Dose Response Experiments In Cancer Research, Biometrika 66 (1979) pp. 307-316

Interagency Regulatory Liaison Group, Scientific Bases For Indentification Of Potential Carcinogens And Estimation Of Risks, Federal Register, July 6, 1979

Krewski, D., Kovar, J., Low Dose Extrapolation Under Single Parameter Dose-Response Models, Comm. In. Stat. B. (1982) 1st Issue

Krewski, D., Kovar, J. and Arnold, D., Optimal Experimental Designs For Low-Dose Extrapolation I, The Case of Zero Background, Submitted to Biometrics (1981-1)

Krewski, D., Kovar, J. and Arnold, D., Optimal Experimental Designs For Low-Dose Extrapolation II, The Case Of Non-Zero Background, Environmetrics 81: Summaries of Conference Presentations, Siam (1981-2) pp. 12-13

Mantel, N. and Bryan, W., "Safety" Testing Of Carcinogenic Agents, J. Nat. Cancer Inst. 27 (1961) pp. 455-470

Mantel, N., Bohidar, N., Brown, C., Ciminera, J. and Tukey, J., An Improved Mantel-Bryan Procedure For "Safety" Testing Of

Carcinogens, Cancer Res. 35 (1975) pp. 865-872

Mirsky, L., Linear Algebra, Oxford University Press (1963)

Naylor, T., Balintfy, J., Burdick, D. and Chu, K., Computer Simulation Techniques, New York, Wiley (1966)

Neyman, J. and Scott, E., Statistical Aspects Of The Problem Of Carcinogenesis, Fifth Berkeley Symposium on Math. Stat. and Prob., Univ. of California Press, Berkeley (1967) pp. 745-776

Poon A., A Monte Carlo Study Of The Power Of Some k-Sample Tests For Ordered Binomial Alternatives, J. Statist. Comput. Simul. 11 (1980) pp. 251-259

Rai, K. and Van Ryzin, J., Risk Assessment Of Toxic Environmental Substances Using A Generalized Multi-Hit Dose-Response Model, Energy and Health, SIAM Press, Phil. (1978)

Schneiderman, M., Mantel, N. and Brown, C., From Mouse To Man - Or How To Get From The Laboratory to Park Avenue and 59th Street, Annals of New York Academy of Sciences 246 (1975) pp. 237-246

Searle, S., Linear Models, New York, Wiley (1971)

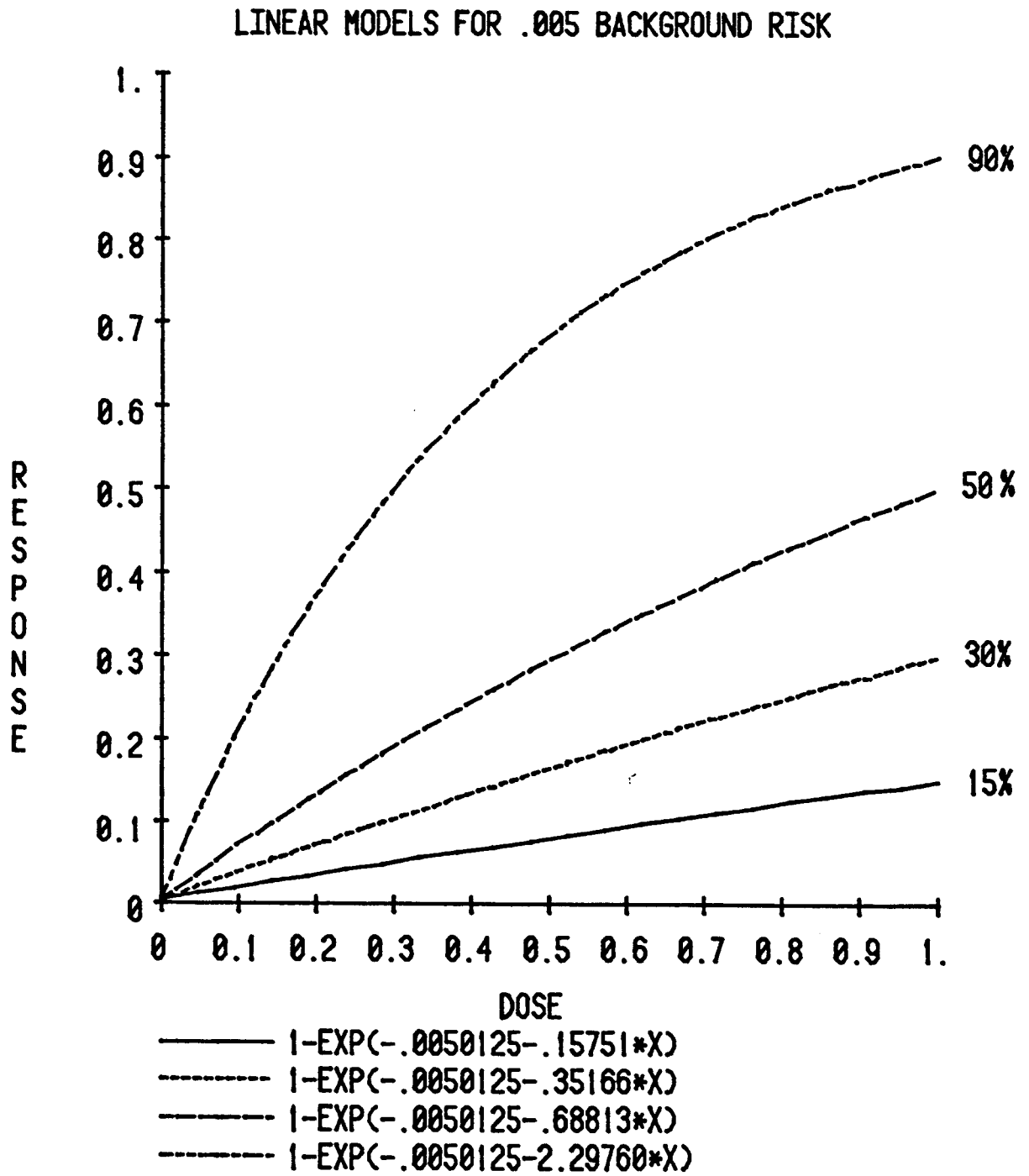
Van Ryzin, J. and Rai, K., The Use Of Quantal Response Data To Make Predictions, The Scientific Basis of Toxicity Assessment, Elsevier/North-Holland, New York (1980) pp. 273-290

Wahrendorf, J., The Problem Of Estimating Safe Dose Levels In Chemical Carcinogenesis, J. Cancer Res. Clin. Oncol. 95 (1979) pp. 101-107

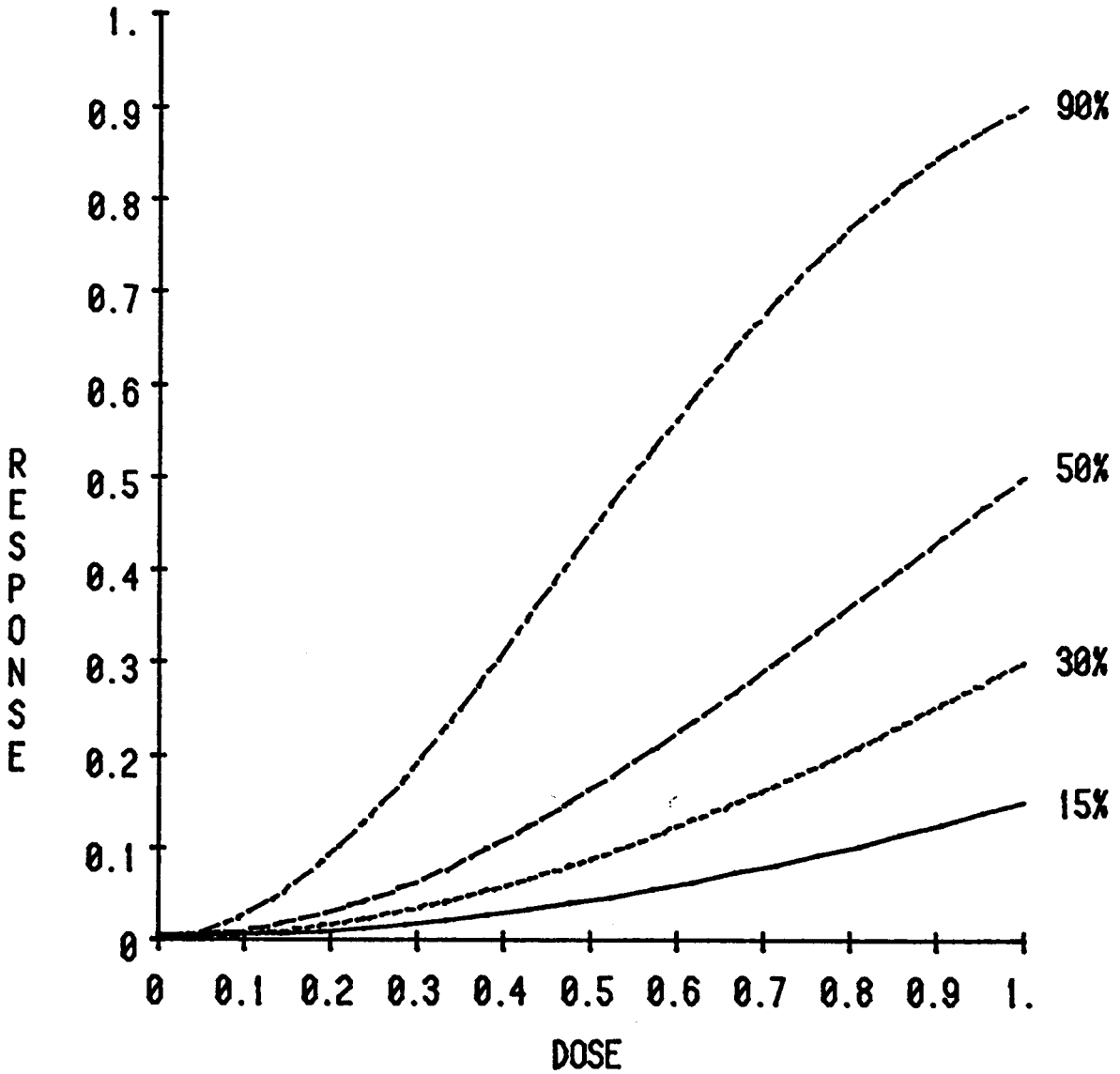
Whittemore, A., Mathematical Models Of Cancer And Their Use In Risk Assessment, Stanford Technical Report No. 27, Dept. of Stat., Standford Univ., Stanford, CA (1979)

Wong, S., Design For Low Dose Extrapolation Of Carcinogenicity Data, Stanford Technical Report No. 48, Dept. of Stat., Standord Univ., Stanford, CA (1979)

APPENDIX A

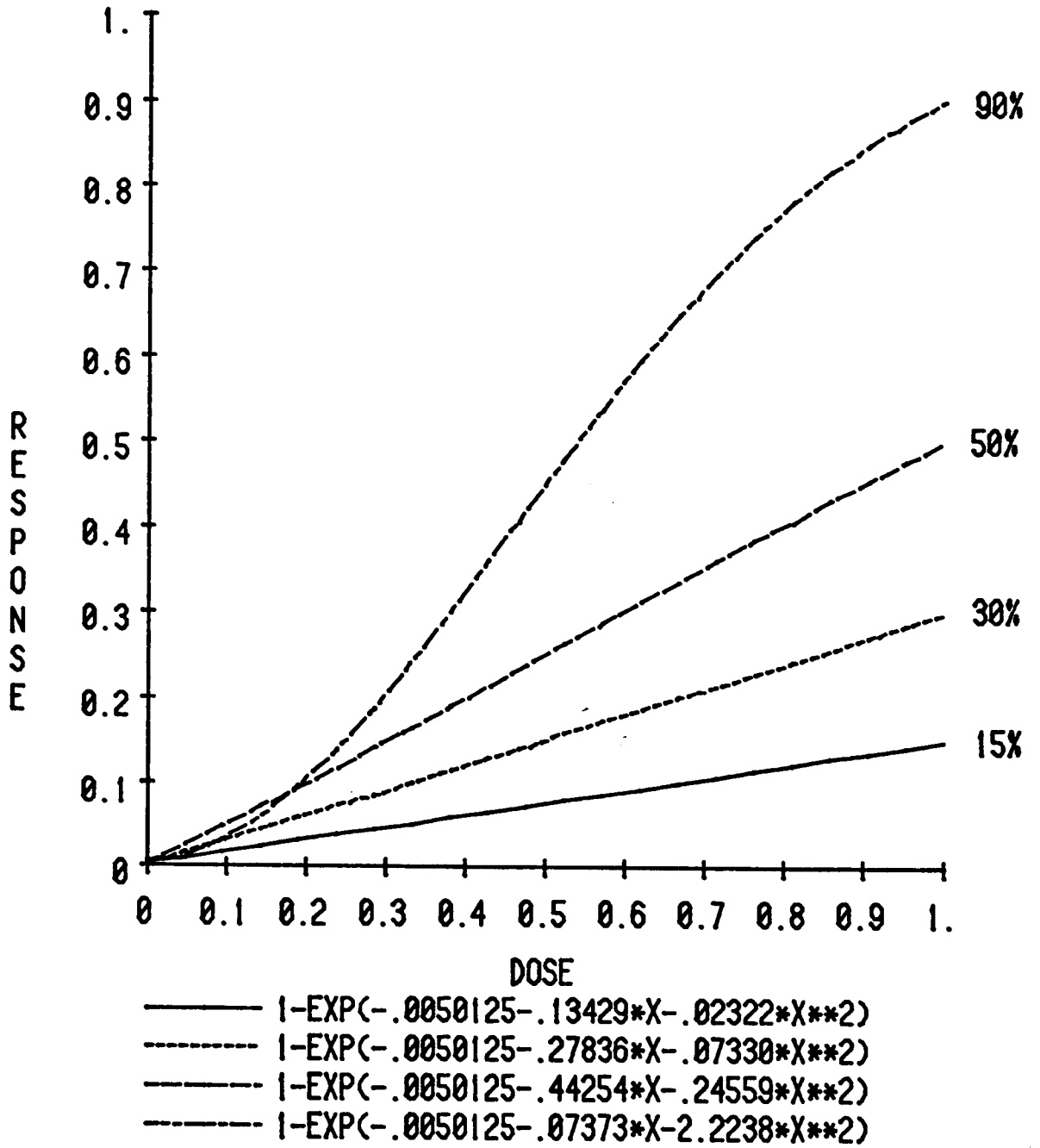

LINEAR MODELS FOR .005 BACKGROUND RISK
RESPONSE
1.
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.
DOSE
90%
50%
30%
15%
1-EXP(-.0050125-.15751*X)
1-EXP(-.0050125-.35166*X)
1-EXP(-.0050125-.68813*X)
1-EXP(-.0050125-2.29760*X)

QUADRATIC MODELS FOR .005 BACKGROUND RISK

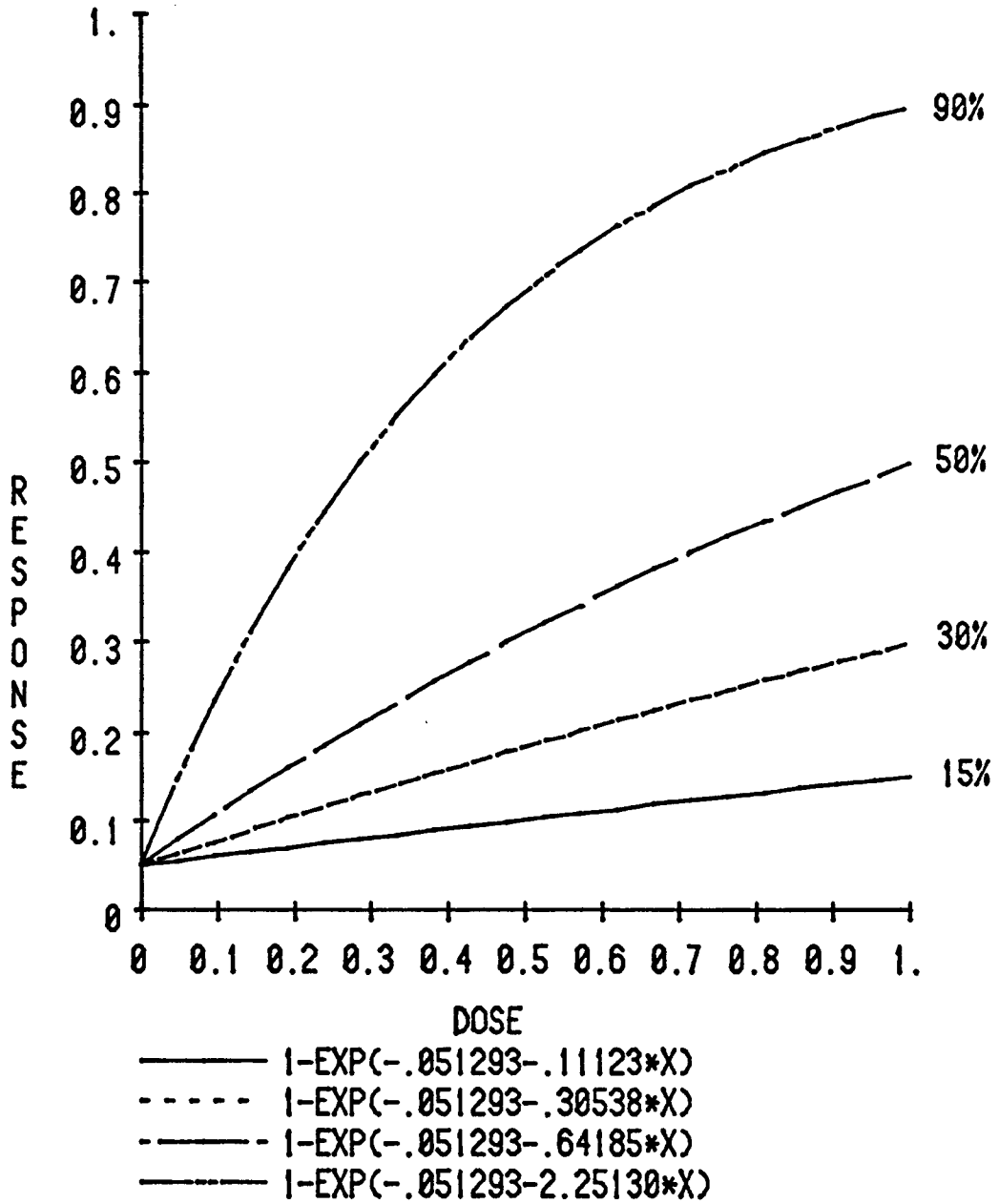


——— 1-EXP(-.0050125-.15751*X**2)
- - - - 1-EXP(-.0050125-.35166*X**2)
—·—·— 1-EXP(-.0050125-.68813*X**2)
—··—··— 1-EXP(-.0050125-2.2976*X**2)

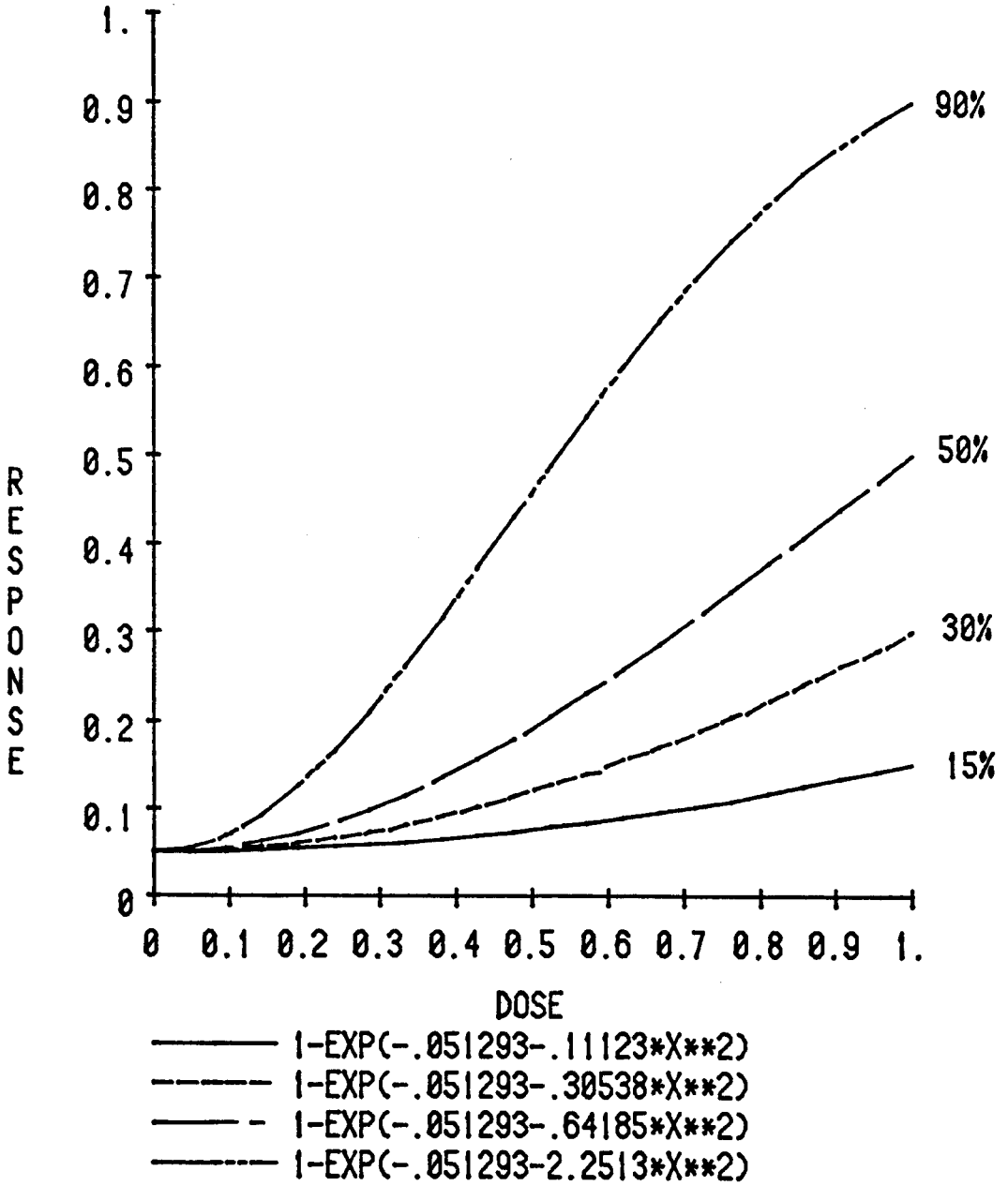
LINEAR-QUADRATIC MODELS FOR .005 BACKGROUND RISK



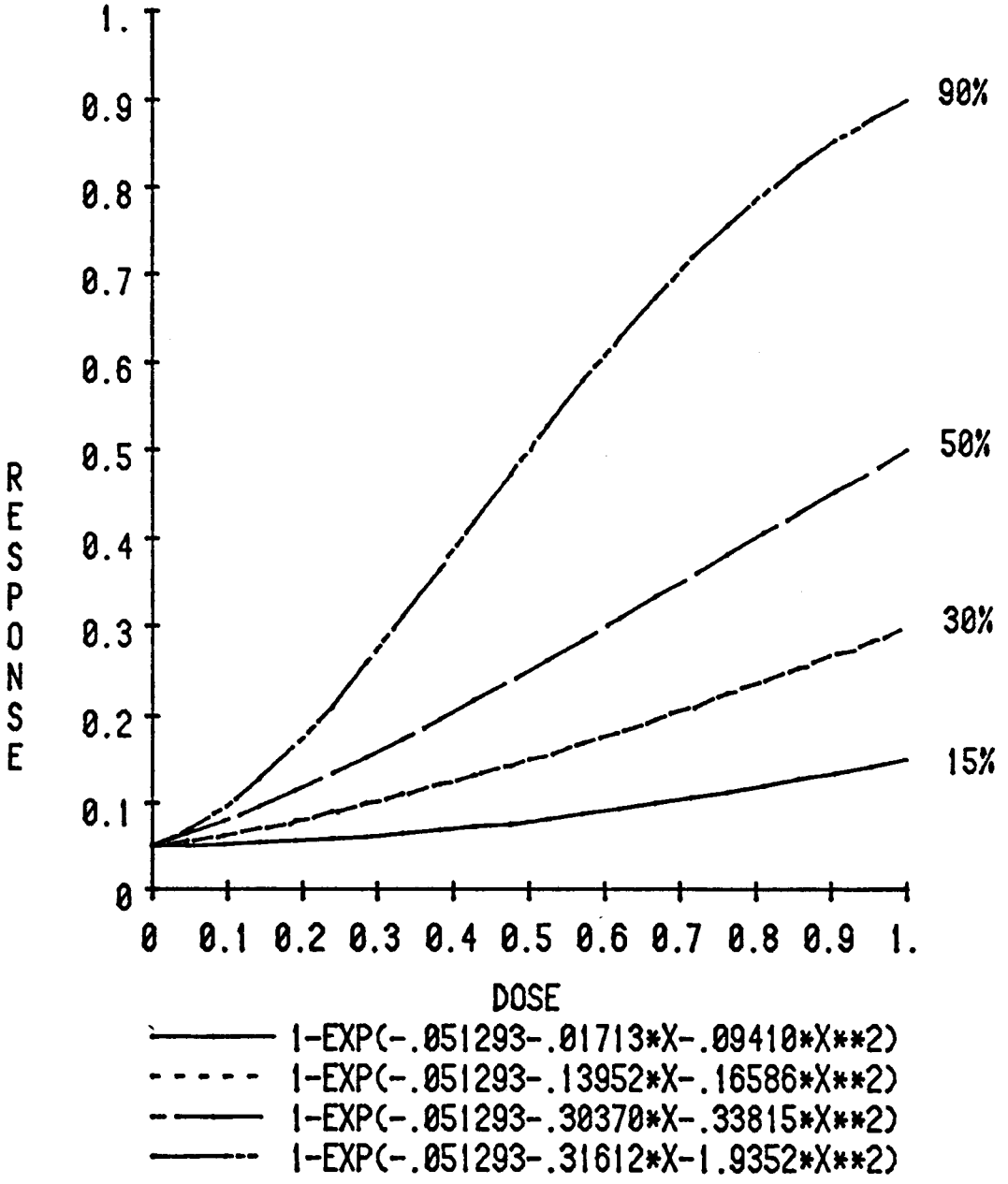
LINEAR MODELS FOR .05 BACKGROUND RISK

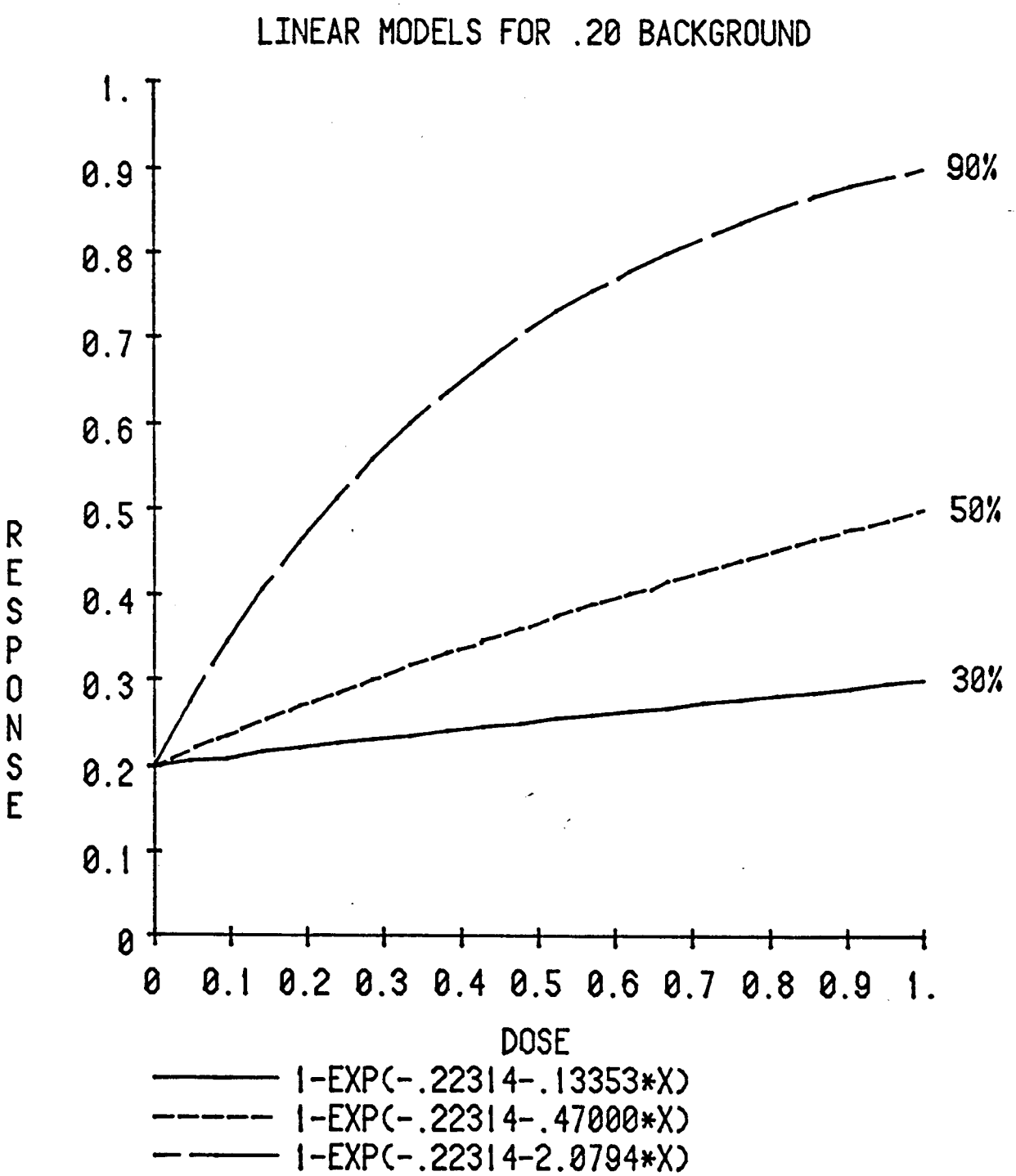


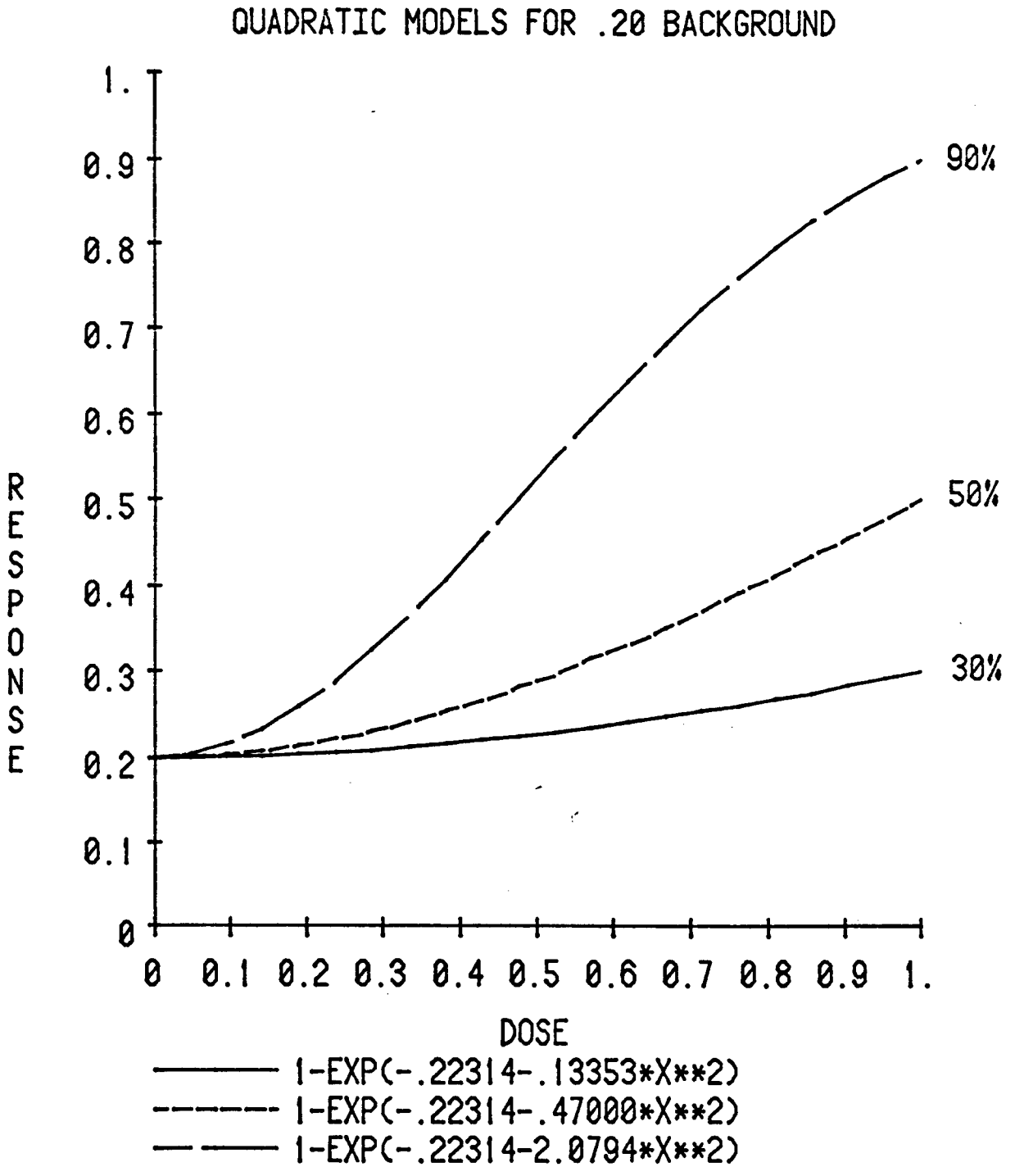
QUADRATIC MODELS FOR .05 BACKGROUND

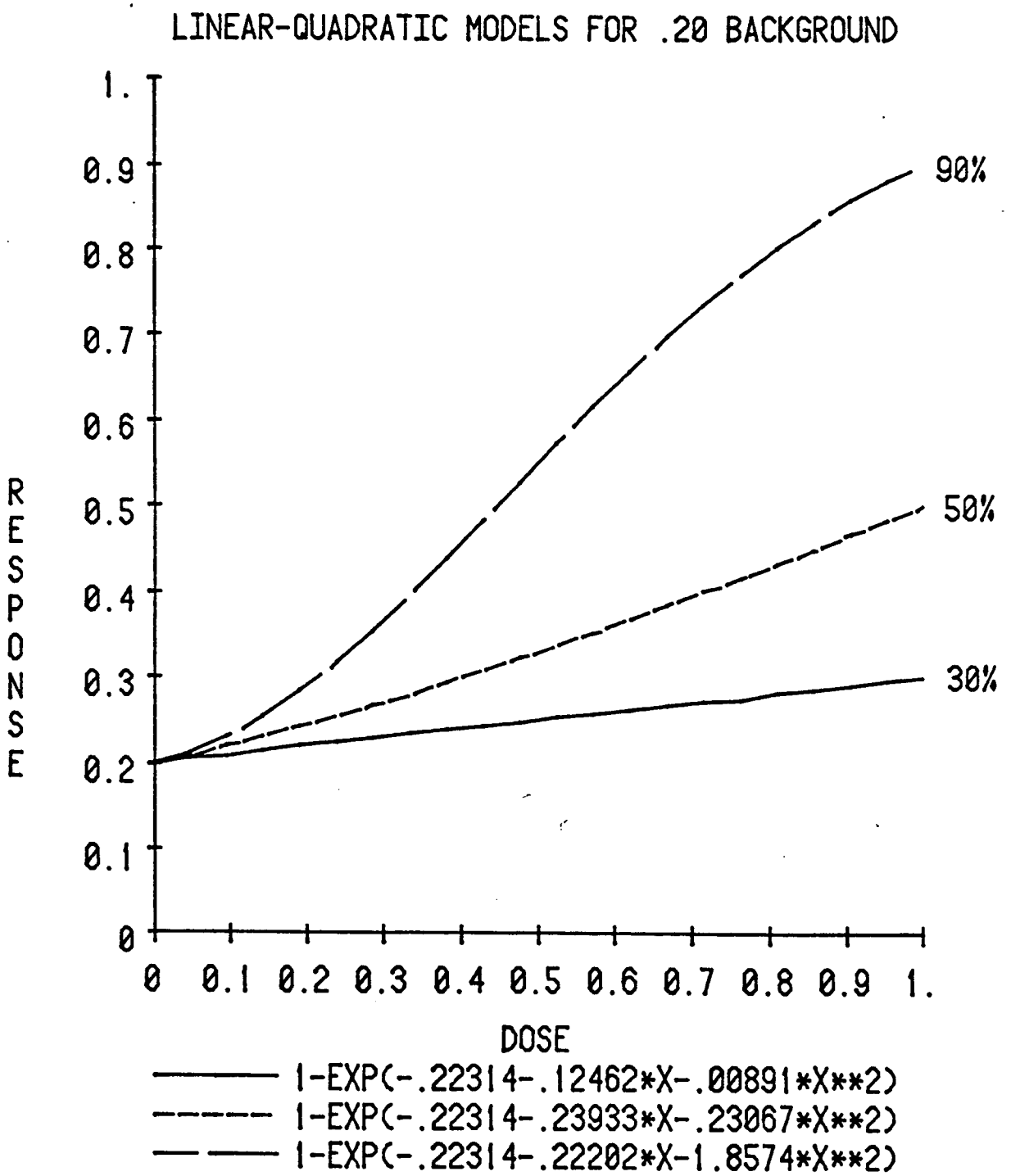


LINEAR-QUADRATIC MODELS FOR .05 BACKGROUND




LINEAR MODELS FOR .20 BACKGROUND
RESPONSE
1.
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
90%
50%
30%
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.
DOSE
1-EXP(-.22314-.13353*X)
1-EXP(-.22314-.47000*X)
1-EXP(-.22314-2.0794*X)


QUADRATIC MODELS FOR .20 BACKGROUND
RESPONSE
1.
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.
DOSE
90%
50%
30%
1-EXP(-.22314-.13353*X**2)
1-EXP(-.22314-.47000*X**2)
1-EXP(-.22314-2.0794*X**2)


LINEAR-QUADRATIC MODELS FOR .20 BACKGROUND
RESPONSE
1.
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
90%
50%
30%
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.
DOSE
1-EXP(-.22314-.12462*X-.00891*X**2)
1-EXP(-.22314-.23933*X-.23067*X**2)
1-EXP(-.22314-.22202*X-1.8574*X**2)

APPENDIX B

Asymptotically Optimal Designs For An Acceptable Added Risk Of 10^{-4}

Back-ground	Model Type	MTD Response	AO Designs d_0	d_1	d_2	γ_0	γ_1	γ_2
.005	LINEAR	.15	0.	.353	1.	.180	.647	.173
		.30	0.	.328	1.	.142	.685	.173
		.50	0.	.306	1.	.117	.710	.173
		.90	0.	.241	1.	.090	.742	.168
	QUADRAT	.15	0.	.271	1.	.295	.577	.128
		.30	0.	.225	1.	.269	.606	.125
		.50	0.	.189	1.	.250	.627	.123
		.90	0.	.125	1.	.220	.640	.140
	LIN-QUA	.15	0.	.315	1.	.198	.677	.125
		.30	0.	.284	1.	.161	.716	.123
		.50	0.	.263	1.	.142	.736	.122
		.90	0.	.134	1.	.195	.657	.148
.050	LINEAR	.15	0.	.448	1.	.322	.512	.166
		.30	0.	.401	1.	.285	.546	.169
		.50	0.	.367	1.	.255	.574	.171
		.90	0.	.280	1.	.218	.613	.169
	QUADRAT	.15	0.	.421	1.	.358	.508	.134
		.30	0.	.364	1.	.338	.521	.141
		.50	0.	.313	1.	.320	.534	.146
		.90	0.	.216	1.	.287	.535	.178
	LIN-QUA	.15	0.	.422	1.	.349	.514	.137
		.30	0.	.371	1.	.311	.553	.136
		.50	0.	.327	1.	.283	.580	.137
		.90	0.	.225	1.	.262	.565	.173
.200	LINEAR	.30	0.	.484	1.	.361	.481	.158
		.50	0.	.438	1.	.344	.493	.163
		.90	0.	.331	1.	.322	.511	.167
	QUADRAT	.30	0.	.466	1.	.371	.501	.128
		.50	0.	.412	1.	.355	.504	.141
		.90	0.	.296	1.	.323	.496	.181
	LIN-QUA	.30	0.	.466	1.	.363	.510	.127
		.50	0.	.414	1.	.348	.518	.134
		.90	0.	.297	1.	.320	.504	.176

Asymptotically Optimal Designs For An Acceptable Added Risk Of 10^{-8}

Back-ground	Model Type	MTD Response	AO Designs d_0	d_1	d_2	γ_0	γ_1	γ_2
.005	LINEAR	.15	0.	.353	1.	.179	.648	.173
		.30	0.	.328	1.	.142	.685	.173
		.50	0.	.306	1.	.117	.710	.173
		.90	0.	.241	1.	.090	.742	.168
	QUADRAT	.15	0.	.270	1.	.293	.571	.136
		.30	0.	.225	1.	.267	.602	.131
		.50	0.	.189	1.	.247	.624	.129
		.90	0.	.125	1.	.218	.636	.146
	LIN-QUA	.15	0.	.315	1.	.198	.677	.125
		.30	0.	.284	1.	.161	.716	.123
		.50	0.	.263	1.	.142	.736	.122
		.90	0.	.134	1.	.195	.656	.149
.050	LINEAR	.15	0.	.448	1.	.322	.512	.166
		.30	0.	.401	1.	.285	.546	.169
		.50	0.	.367	1.	.255	.574	.171
		.90	0.	.280	1.	.218	.613	.169
	QUADRAT	.15	0.	.421	1.	.354	.507	.139
		.30	0.	.363	1.	.336	.519	.145
		.50	0.	.313	1.	.318	.532	.150
		.90	0.	.216	1.	.284	.534	.182
	LIN-QUA	.15	0.	.422	1.	.349	.514	.138
		.30	0.	.371	1.	.311	.553	.136
		.50	0.	.327	1.	.283	.580	.137
		.90	0.	.224	1.	.264	.564	.172
.200	LINEAR	.30	0.	.484	1.	.361	.481	.158
		.50	0.	.438	1.	.344	.493	.163
		.90	0.	.331	1.	.322	.511	.167
	QUADRAT	.30	0.	.465	1.	.367	.501	.132
		.50	0.	.412	1.	.353	.503	.144
		.90	0.	.296	1.	.320	.496	.184
	LIN-QUA	.30	0.	.466	1.	.364	.509	.127
		.50	0.	.414	1.	.348	.518	.134
		.90	0.	.297	1.	.320	.504	.176

APPENDIX C

ASYMPTOTIC RELATIVE EFFICIENCY TO THE OPTIMAL DESIGN

LOW-DOSE=	0.00	0.00	0.00	0.00	0.00
MID-DOSE=	0.20	0.30	0.40	0.50	0.60
HIGH-DOSE=	1.00	1.00	1.00	1.00	1.00
% AT LOW =	0.10	0.10	0.10	0.10	0.10
% AT MID =	0.60	0.60	0.60	0.60	0.60
% AT HIGH=	0.30	0.30	0.30	0.30	0.30
MODELS:					
LI005M15	0.64	0.86	0.92	0.82	0.62
LI005M30	0.75	0.92	0.92	0.78	0.56
LI005M50	0.82	0.95	0.90	0.73	0.50
LI005M90	0.91	0.92	0.76	0.54	0.32
QD005M15	0.50	0.72	0.73	0.58	0.37
QD005M30	0.66	0.79	0.68	0.47	0.27
QD005M50	0.78	0.78	0.58	0.36	0.20
QD005M90	0.83	0.56	0.32	0.17	0.08
LQ005M15	0.58	0.81	0.85	0.73	0.52
LQ005M30	0.70	0.87	0.84	0.67	0.44
LQ005M50	0.79	0.89	0.78	0.58	0.36
LQ005M90	0.87	0.60	0.35	0.19	0.09
LI050M15	0.27	0.46	0.61	0.69	0.67
LI050M30	0.38	0.60	0.73	0.76	0.67
LI050M50	0.50	0.71	0.81	0.78	0.63
LI050M90	0.74	0.87	0.82	0.65	0.43
QD050M15	0.21	0.38	0.54	0.63	0.61
QD050M30	0.28	0.49	0.65	0.67	0.55
QD050M50	0.37	0.61	0.72	0.64	0.44
QD050M90	0.67	0.78	0.60	0.36	0.19
LQ050M15	0.21	0.39	0.55	0.64	0.62
LQ050M30	0.31	0.53	0.68	0.71	0.59
LQ050M50	0.42	0.66	0.76	0.70	0.51
LQ050M90	0.69	0.82	0.65	0.41	0.22
LI200M30	0.20	0.36	0.51	0.60	0.63
LI200M50	0.28	0.46	0.59	0.66	0.63
LI200M90	0.48	0.64	0.69	0.65	0.51
QD200M30	0.17	0.32	0.46	0.57	0.61
QD200M50	0.22	0.40	0.56	0.64	0.60
QD200M90	0.40	0.65	0.71	0.57	0.35
LQ200M30	0.18	0.33	0.48	0.58	0.62
LQ200M50	0.23	0.41	0.57	0.65	0.61
LQ200M90	0.41	0.65	0.71	0.58	0.37

ASYMPTOTIC RELATIVE EFFICIENCY TO THE OPTIMAL DESIGN

LOW-DOSE=	0.00	0.00	0.00	0.00	0.00
MID-DOSE=	0.20	0.30	0.40	0.50	0.60
HIGH-DOSE=	1.00	1.00	1.00	1.00	1.00
% AT LOW =	0.30	0.30	0.30	0.30	0.30
% AT MID =	0.60	0.60	0.60	0.60	0.60
% AT HIGH=	0.10	0.10	0.10	0.10	0.10
MODELS:					
LI005M15	0.76	0.92	0.87	0.67	0.42
LI005M30	0.79	0.89	0.80	0.59	0.36
LI005M50	0.79	0.86	0.74	0.52	0.31
LI005M90	0.82	0.75	0.54	0.33	0.18
QD005M15	0.89	0.96	0.71	0.42	0.22
QD005M30	0.99	0.84	0.54	0.30	0.15
QD005M50	0.97	0.69	0.40	0.21	0.11
QD005M90	0.68	0.34	0.16	0.08	0.04
LQ005M15	0.82	0.96	0.83	0.58	0.34
LQ005M30	0.85	0.90	0.73	0.48	0.28
LQ005M50	0.88	0.84	0.62	0.39	0.21
LQ005M90	0.71	0.37	0.18	0.09	0.04
LI050M15	0.51	0.79	0.95	0.91	0.70
LI050M30	0.63	0.89	0.97	0.84	0.58
LI050M50	0.73	0.94	0.93	0.74	0.47
LI050M90	0.91	0.93	0.73	0.47	0.26
QD050M15	0.47	0.80	0.96	0.88	0.61
QD050M30	0.61	0.92	0.94	0.70	0.42
QD050M50	0.76	0.97	0.81	0.52	0.28
QD050M90	0.95	0.68	0.37	0.18	0.09
LQ050M15	0.48	0.80	0.97	0.88	0.62
LQ050M30	0.63	0.93	0.96	0.75	0.47
LQ050M50	0.77	0.99	0.88	0.60	0.35
LQ050M90	0.96	0.73	0.42	0.21	0.10
LI200M30	0.43	0.70	0.89	0.94	0.80
LI200M50	0.53	0.80	0.94	0.90	0.69
LI200M90	0.78	0.96	0.88	0.65	0.39
QD200M30	0.39	0.69	0.91	0.95	0.76
QD200M50	0.49	0.82	0.96	0.84	0.56
QD200M90	0.79	0.92	0.66	0.37	0.18
LQ200M30	0.41	0.71	0.92	0.95	0.77
LQ200M50	0.51	0.83	0.97	0.87	0.60
LQ200M90	0.80	0.93	0.68	0.39	0.20