DESIGN FOR LOW DOSE EXTRAPOLATION OF CARCINOGENICITY DATA

BY

SUZANNA CHOR-YEE WONG

TECHNICAL REPORT NO. 24

JUNE 1979

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH (SIMS)

PREPARED UNDER SUPPORT TO SIMS FROM

DEPARTMENT OF ENERGY (DOE)

ROCKEFELLER FOUNDATION

SLOAN FOUNDATION

ENVIRONMENTAL PROTECTION AGENCY (EPA)

NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *JB*

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ACKNOWLEDGMENTS

I deeply thank Professor Lincoln E. Moses, my dissertation advisor, for his invaluable suggestions, encouragement, guidance and assistance.

Special thanks go to Professor Paul Switzer and Professor Byron William Brown, Jr., my reading committee members, for their very helpful suggestions towards my dissertation.

I am also very grateful to the many people who helped and guided my way in my graduate studies. I would like to express my appreciation to all my professors for the education I received, to the SIMS group and the Statistics Department for their financial support, to Ms Judi Davis for her arrangement in the typing of this dissertation, and to my dearest parents and my husband for their full support and encouragement.

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
	1.1. The Problem of Low Dose Extrapolation.....	1
	1.2. Experimental Setup	2
	1.3. Review of Models	4
	1.4. Statement of Problem and Outline	7
II	DESIGN FOR LARGE SAMPLE LOW DOSE ESTIMATION.....	10
	2.1. Model for Estimation	10
	2.2. General Approach to the Design Problem.....	15
	2.3. Design for Estimating Low Dose Response Rate	25
	2.4. Design for Estimating Safe Dose	47
III	A SMALL SAMPLE INVESTIGATION OF THE USE OF DIFFERENT BERNOULLI ESTIMATORS FOR LOW DOSE EXTRAPOLATION	60
	3.1. Motivation	60
	3.2. The One Parameter Case	63
	3.3. The Two Parameter Case	88
IV	DIFFERENTIATION OF MODELS WITH DIFFERENT LOW DOSE BEHAVIORS	103
	4.1. Approach to the Model Testing Problem	103
	4.2. Design for Testing Different Low Dose Behaviors	110
	4.3. Conclusion	135
	REFERENCES	142

CHAPTER I

INTRODUCTION

1.1. The Problem of Low Dose Extrapolation

Many substances in our environment, e.g., air pollutants, food additives, radiations, and chemicals, are suspected cancer-causing agents. They exist in the environment at such low levels that their carcinogenicity may necessarily be obscure; yet on experimentation, animals exposed to them at higher levels may develop cancer. Studies are thus needed to assess carcinogenicity.

Studies on environmental carcinogens are often aimed at answering two questions. For a given carcinogen:

1. What is the risk to the human population at its existing level of exposure?
2. What is the highest safe exposure level for human?

Answers to these questions are important for control purposes, and since the existing and the safe exposure levels of carcinogens are low, both require inferences on the low-exposure effects.

Carcinogenicity data come from two sources: epidemiologic studies or laboratory studies. Epidemiologic studies are made directly on the human population, but are complicated by many interfering factors. Individuals are exposed simultaneously to several carcinogens and the exposure levels, which vary tremendously with time and individual, are usually poorly documented. The long delay between the initial

exposure to a carcinogen and the appearance of cancer in an individual makes accurate experimentation even more difficult, and as a result, conclusions from epidemiologic studies are often weak.

Laboratory studies have two main drawbacks: the number of experimental subjects is limited and animals have to be used in the experiments. The first limitation implies that experimentation cannot be performed at the low dose range, the region where inference is to be drawn, but has to be restricted to the middle dose range, so that at least some, but not all, of the subjects put to test will develop cancer. The second limitation implies that experimental conclusions on animals at low doses require reinterpretation for applicability to human. The first problem, called the low dose extrapolation problem, is considered here. The main concern is to draw inferences on the exposure effect at the low dose range from observations in the relatively high dose range.

1.2. Experimental Setup

A typical laboratory setup has several groups of experimental subjects exposed continuously to fixed but different levels of a given carcinogen. The effect of exposure is to be identified by a response, to be defined, e.g., as death due to a tumor, and the subjects are observed for their responses.

Different kinds of data may be obtained from the experiment. If the response is, e.g., the clinical appearance of skin tumor, the time-to-response for each subject is observable and the mean time-to-

response measures the exposure effect. If the response is, e.g., the appearance of a stomach tumor, the time-to-response is not observable. In such cases, only dichotomous data, the presence or absence of the response in a given period of time, for the subjects can be obtained. The probability of a subject responding within the given time period is then the measure of the exposure effect. Since time-to-response is generally not observable and usually not well observed, experiments with dichotomous data are common in carcinogenicity testing; only dichotomous data will be considered here. Further, since each stage of an experiment of this type will normally take a few months to years to complete, we will restrict our attention to a two-stage experiment.

The first stage of our experiment is a preliminary experiment with the above setup. The experimental period is set to be the average life time of the animals used. The purpose of this first stage experiment is to provide some, though not necessarily very precise, information on the probabilities of response at various levels of exposure. The preliminary experiment will not be discussed any further, but such information is assumed available for the second stage of our experiment.

The second stage is the primary experiment for discussion: its results are to be used for inferences on the low-exposure effects. The length of the experimental period is the same as that in the first stage; a substantial number of subjects will be used, and the dose levels are all set in the observable range. Assuming the number of subjects available for this part of experimentation is of the order

of 10^3 , we will define the observable range as the dose range with probability of response between .001 and .999.

1.3. Review of Models

The data from the experiment is used to explore the relationship between the exposure effect and the dose level of the test carcinogen. Such a relationship can often be represented by a parametric model, but in this problem, since inference is to be drawn outside the experimental range, the assumption of a parametric model is a necessity.

Models for dichotomous data are called dichotomous response models. The assumption is the existence of a deterministic dose response function which is known except for a certain number of parameters:

$$p(d) = \text{Pr}(\text{a subject exposed to a dose } d \text{ will respond within a given time period}), \quad 0 \leq d < \infty.$$

We will assume $p(d) = 0$ at $d = 0$ for all the models.

Three classes of dichotomous response models have been discussed in the literatures [1], [3], [6], [7], [8], [11], [12], [14].

1. The threshold model

The model assumes:

$$\begin{aligned} p(d) &= 0, & 0 \leq d \leq d_t \\ &> 0, & d_t < d < \infty. \end{aligned}$$

The threshold d_t is the parameter of interest. Models with $p(d) = 0$ only at $d = 0$ are more popular because statistically it is impossible to prove the existence of a threshold, and controlling the level of a carcinogen to its threshold, even if it were known, may be too stringent a requirement.

2. The tolerance distribution models

This class of models assumes the existence of a tolerance level for every subject in the population. The tolerance level T is the dose level below which a subject will not respond and above which the subject will respond. Except for the unknown location and scale parameters, $\ln T$ is assumed to have a known distribution. Hence a model can be specified by an arbitrary distribution function F :

$$p(d) = F(\alpha + \beta \ln d), \quad -\infty < \alpha < \infty, \quad 0 < \beta < \infty, \quad \alpha, \beta \text{ unknown.}$$

The probit and logistic models in this group have long been used in bioassays, but there is little justification for their being used for extrapolation. Other models include the extreme value distribution model, the rectangular distribution model, and the angular distribution model. The hit models may also be regarded as belonging to this category (with β a known integer) but will be introduced in the next group.

3. The hit models and their extensions

The k-hit model assumes:

$$p(d) = 1 - e^{-\theta d^k}, \quad 1 \leq k < \infty \quad \text{known integer,} \\ 0 < \theta < \infty \quad \text{unknown,}$$

the multistage model assumes:

$$p(d) = 1 - \exp\left(-\sum_{i=1}^k \theta_i d^i\right), \quad 1 \leq k < \infty \quad \text{known integer,} \\ 0 \leq \theta_i < \infty \quad \text{unknown, not all zero,}$$

and the Guess and Crump model assumes:

$$p(d) = 1 - \exp\left(-\sum_{i=1}^k \theta_i d^i\right), \quad 0 \leq \theta_i < \infty \quad \text{unknown,} \\ \text{not all zero,} \\ 1 \leq k < \infty \quad \text{unknown integer.}$$

This class of models is deduced from hypotheses on carcinogenic processes. The k-hit model supposes that a target cell has to receive k "hits" of the biologically effective agent to start an irreversible change leading to the response. The multistage model hypothesizes that carcinogenesis is an irreversible self-replicating k-stage process with some stages depending on the dose of the given carcinogen and the others having spontaneous occurrences. The Guess and Crump model allows the number of stages k to be free. Validity of these hypotheses are still unknown with the present knowledge on the mechanisms of carcinogenic processes.

Further discussion of dichotomous response models appears in Chapters II and IV.

There is also a second class of models known as the time-to-response models for experiments with time-to-response data. Two models are commonly used: the weibull model assumes the time-to-response has a weibull distribution and the log-normal model assumes a log-normal distribution. Since we will work with dichotomous data, we will not discuss these models any further, except to remark that they are related to the dichotomous response models [3], namely, the extreme value distribution and the probit models respectively.

1.4. Statement of Problem and Outline

The formulation for the low dose extrapolation problem is as follows:

1. The experimental procedure

The experiment is as described in Section 1.2; the following notation is used for the specification of the second stage of the experiment:

N = the total number of experimental subjects,

g = the number of experimental groups,

n_i = the number of subjects in group i , $i = 1, \dots, g$,

i.e. $\sum_{i=1}^g n_i = N$,

d_i = the dose level of the test carcinogen for group i ,

$i = 1, \dots, g$,

the dose levels are different for different groups,

b_1, b_2 are the lower and upper boundaries of the observable dose range, they are only approximately known and

$$0 < b_1 \leq d_i \leq b_2, i = 1, \dots, g,$$

t = the length of the experimental period,

U_{e_j} = the dichotomous response within time t of subject j exposed to a dose e_j , $j = 1, \dots, N$,

X_i = the number of subjects in group i responding within time t , $i = 1, \dots, g$, $X_i = \sum_{j=1}^N U_{e_j} I_{[e_j=d_i]}$.

2. The inferences

Let $p^0(d)$ be the true dose response function, i.e.

$$p^0(d) = \Pr(\text{a subject at dose } d \text{ will respond in time } t), 0 \leq d < \infty.$$

(1) Given a fixed low dose d_0 , say $0 < d_0 < b_1$, inference is to be drawn for the extrapolated low dose response rate $p^0(d_0)$,

(2) given a fixed low response rate p_0 , say $0 < p_0 < 10^{-3}$, inference is to be drawn for the safe dose $d^0(p_0)$ defined by $p^0(d^0(p_0)) = p_0$.

With a model $P = \{p_{\underline{\theta}}(d), 0 \leq d < \infty; \underline{\theta} \in \Theta\}$, i.e. with the assumption $p^0(d) \in P$, the object is to use data

$\{U_{e_j} \sim b(1, p^0(e_j)), j = 1, \dots, N\}$ to identify $p^0(d)$

from the functions in P for making inferences on the low dose behavior.

In several occasions, we will use different parametrizations of the same model, e.g. $P = \{p_{\underline{\theta}}(d), 0 \leq d < \infty; \underline{\theta} \in \Theta\}$ $= \{p_{\underline{\zeta}}(d), 0 \leq d < \infty; \underline{\zeta} \in Z\}$. We will use the same symbol P for the two classes of functions and refer to the dose response functions as $p(d)$ when we do not want to make the distinction on the parametric form used, or we may want to focus at the response rate at a dose d_1 , $p(d_1)$, without referring to the parameters of the model. We will refer to $p_{\underline{\theta}}(d)$, $p_{\underline{\zeta}}(d)$ as functions of d with those particular parametrizations. Hence $p^0(d)$ is the true dose response function corresponding to $p_{\underline{\theta}^0}(d)$, $p_{\underline{\zeta}^0}(d)$, and $\underline{\theta}^0$, $\underline{\zeta}^0$ are the true values of the parameters of those particular parametrizations.

The effectiveness (with respect to a certain criterion) of the inferences depend on three components: the experimental design, the estimation procedure for the inference, and the model assumption. We will discuss these problems in Chapters II to IV. Chapter II discusses the experimental design problem for a proposed model using maximum likelihood (with or without modification) for inference. The number of experimental groups and allocations of sample sizes and dose levels are determined for purposes of estimating the low dose response rate and the safe dose. Chapter III investigates, for small sample sizes, the relation between the design space and methods of estimation for low dose extrapolation for two cases of the proposed model. Chapter IV proposes a more general model which includes the model in Chapter II as a subclass and investigates the design for and the possibility of detecting differences between models with different low dose behaviors.

CHAPTER II

DESIGN FOR LARGE SAMPLE LOW DOSE ESTIMATION.

2.1. Model for Estimation

The dichotomous response models commonly used for low dose extrapolation have been reviewed in Section 1.3. On grounds of the present limited knowledge on the mechanisms of carcinogenic processes, no comparisons of the biological justifications of such models will be made, but instead, we will look at these models and evaluate them from a purely mathematical point of view. We will first review the literature on the justifications and problems of the different models.

The problem with the tolerance distribution models is the difference in their low dose behaviors. Each of the models exhibits similar response curves between the 10% and the 90% range, but since the distributions vary greatly in their tail weights, the different models fitted to the same data set may look very similar in the experimental range but differ markedly (by as much as several orders of magnitude) in their low dose extrapolated results [3], [12]. This heavy dependence of the low dose behavior on the model makes it very difficult to choose between these models.

The difficulties are of several kinds. There is a lack of biological justification for the use of any of the distributions. It is difficult to differentiate between the models (a huge number of observations is needed to detect the difference between the logistic

and the probit models [2]). There is no guarantee that the best model describing the data is best for extrapolation, and there is no adequate experimental data to suggest the shape of $p^0(d)$ at low doses.

The hit model extensions, the multistage model and the Guess and Crump model, give a new approach to solving the model dependency problem. The idea is to include enough functions in the model so that $p^0(d)$, with whatever kind of possible low dose behavior, is likely to be included in the model, and instead of having to choose a model then a function from the chosen model, the data can be used directly to pick out the true dose response function from the larger model. Two criteria are thus necessary: functions included in the model should:

1. provide adequate representation of the low dose behavior,
2. be reasonably flexible and general.

With no data on the shape of $p^0(d)$ at low doses, we will have to rely on theoretical conjectures. In general, it is agreed that $p^0(d)$ should be concave in d for small $d > 0$. Several arguments support the linear low dose behavior [6], [8], [13] (i.e. support $p^0(d) - p^0(0) = \alpha d$ for small $d > 0$) which is the most conservative for low dose extrapolation (in the sense of having the highest extrapolated response rate and the lowest safe dose among all dose response functions concave in d for small $d > 0$). The argument is by Taylor series expansion of $p^0(d)$ near $d = 0$:

$$p^0(d) - p^0(0) = p^{0(1)}(0)d + \frac{p^{0(2)}(0)}{2!}d^2 + \dots$$

and continues to argue that $p^{(1)}(0) > 0$. However, it seems that, even if the low dose behavior is not linear, it should be sublinear because

$$p^0(d) - p^0(0) \doteq \frac{p^{(l)}(0) d^l}{l!},$$

for small $d > 0$, for some $l = 1, 2, \dots$.

With this in mind, the class of hit models are good representatives of different low dose behaviors:

1. for the k-hit model,

$$p(d) - p(0) \doteq \theta d^k, \quad \text{for small } d > 0,$$

so $p^0(d)$ should behave like one of the hit models at low doses;

2. there is a wide range of low dose behaviors: the functions flatten from a linear function to an almost threshold like function at low doses as k increases;
3. the 1-hit model is linear at low doses.

The Guess and Crump model essentially joins the class of hit models through the exponential of a non-negative coefficient polynomial of arbitrary degree, thus all the hit models are included to adequately represent the low dose behavior. The functional form also adds flexibility and generality to the model. It allows for a different high dose behavior as the higher order terms become more dominant at higher doses and its generality is reflected in the completeness

of the system of non-negative coefficient, polynomial hazard-rate-functions in the class of positive continuous hazard-rate-functions [11]. The multistage model has similar properties but to a lesser extent.

The Guess and Crump model has been used to analyze real and hypothetical data sets. The data sets on chemicals and radiations analyzed by the constrained maximum likelihood method [8], [9], [10] all yield positive estimates to θ_1 indicating that the low dose extrapolation may be closer to linearity than a flat curve like the probit. Analyses of hypothetical data sets reveal some difficulties [7], [8]:

1. it is difficult to reject the linear model even if $p^0(d)$ is highly non-linear and large samples are used for the experiment;
2. $p^0(d)$ with a small linear term and some higher order terms may result with an estimated dose response function without the linear term, extrapolations from the estimated function and from $p^0(d)$ are thus very different;
3. in an experiment with eight dose levels and a thousand subjects per dose level: two very similar hypothetical data sets (with responses at three of the dose levels differing in numbers small compared with their standard deviations) may result in a radical change in extrapolations from their estimated response functions.

We interpret the first two results as indications of the difficulties in estimation and testing of tail behaviors. We consider the last result a very disturbing one, however, and reason that it comes from allowing the choice of the low dose behavior to be made

from a discrete set of behaviors (i.e. the exponent of d can be 1, 2, 3, etc.). Our considerations have led us to investigate the extreme value distribution model for estimations in this chapter:

$$p(d) = 1 - \exp(-e^{\alpha+\beta \ln d}), \quad -\infty < \alpha < \infty, 0 < \beta < \infty$$

$$= 1 - e^{-\theta d^\beta}, \quad 0 < \theta < \infty, 0 < \beta < \infty.$$

The reasons are:

1. the extreme value distribution model has curves similar to the other tolerance distribution models in the observable range so that the fitness to the data in the observable range may be considered as adequate;
2. it includes all the hit models for adequate representation of low dose behaviors;
3. it allows for continuous graduation of dose response functions between those which are interpretable as "hit models."

The disadvantage is the lack of flexibility (that of assuming the same high and low dose behaviors), so that the model dependency problem has to be investigated. We will attempt the problem of detecting different low dose behaviors in Chapter IV.

2.2. General Approach to the Design Problem

Designs for estimating the low dose response rate and the safe dose are investigated in this chapter. The second stage of our experiment, as described in Sections 1.2 and 1.4, presupposes the availability of:

1. a large number of subjects for this part of the experimentation, and
2. a preliminary experiment which gives information on the dose response function, i.e., if a model $P = \{p_{\underline{\zeta}}(d), 0 \leq d < \infty; \underline{\zeta} \in Z\}$ is assumed, the preliminary experiment will give approximate value to the parameter $\underline{\zeta}$ of the model.

As a consequence, we will search for asymptotic locally optimal designs for this second stage of experimentation (i.e. designs which are optimal if the parameters are known to be close to certain specified values). We will describe the general approach in this section and consider the specific problems of estimating the low dose response rate and the safe dose in the next two sections.

Chernoff gave the asymptotic locally optimal design for estimating parameters in a very general setting [4]. He dealt with the following situation: a set of pure experiments $\mathcal{E} = \{E\}$ is available for experimentation, the probability distribution of data from any $E \in \mathcal{E}$ depends on

$$\underline{\zeta} = \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_k \end{pmatrix} ;$$

the interest is in estimating s ($\leq k$) of the parameters, say

$\xi_1, \dots, \xi_s$, and the design problem is to choose n of the experiments

$E_1, \dots, E_n \in \mathcal{E}$ (possibly with repetitions) to be performed independently.

Let ξ^0 be the true value of the parameter,

$L(\xi)$ be the likelihood function of the data from an experiment

when ξ obtains,

$X(\xi)$ be the Fisher information matrix of the experiment when

ξ obtains, i. e.

$$X(\xi) = \left[E_{\xi} \left(\frac{\partial \ln L(\xi)}{\partial \xi_i} \frac{\partial \ln L(\xi)}{\partial \xi_j} \right) \right]_{i,j=1,\dots,k}$$

For locally optimal design, if ξ is the specified approximate value

of ξ^0 , then $X(\xi)$ is to be examined. Since $X(\xi)$ is a non-negative

definite symmetric matrix, the extended inverse $X^{-1}(\xi) = [x^{ij}(\xi)]$,

$i, j=1, \dots, k$, is defined, and there is a $[k(k+1)]/2$ -dimensional vector

analogue containing the elements on and above the main diagonal of

$X(\xi)$ according to some order.

Let R_1 be the set of such vectors corresponding to $X(\xi)$ for

experiments in $\mathcal{E}$,

R be the convex hull of R_1 representing the set of informa-

tion matrices for the class of mixed experiments of $\mathcal{E}$.

Chernoff's theorem states that if R is closed and bounded, then:

1. there exists an element $\tilde{X} \in R$ which minimizes $v_s(X) = \sum_{i=1}^s x^{ii}(\underline{\xi})$,
2. $\tilde{X}$ is a convex linear combination of $r \leq k + (k-1) + \dots + (k-s+1)$ elements $\tilde{X}_1, \dots, \tilde{X}_r \in R_1$,
3. $\tilde{X}_1, \dots, \tilde{X}_r$ may be chosen so that $v_s(X)$ is a continuous function at $X = \tilde{X}$ with respect to the topology of the convex set generated by $\tilde{X}_1, \dots, \tilde{X}_r$.

The implication of the theorem is that: under the hypothesis, the design minimizing $v_s(X)$ is a mixed experiment which, for large n , may be approximated by selecting a set of r experiments in $\mathcal{E}$ and by repeating each of the r experiments in certain specified proportions. Such a design is locally optimal over all designs and the criterion of optimality has special interpretation [4].

Since our interest is in estimating $p^0(d_0)$ or $d^0(p_0)$, we will use Chernoff's theorem with $s = 1$. Let

$$P = \left\{ p_{\underline{\xi}}(d), 0 \leq d < \infty; \underline{\xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix} \in Z \right\}$$

be our k parameter model containing $p^0(d)$ and such that ξ_1 , say, is the parameter for $p^0(d_0)$ or $d^0(p_0)$. A set of experiments $\mathcal{E} = \{E_d: b_1 \leq d \leq b_2\}$ are hence available, E_d being the experiment of observing the dichotomous response $U_d \sim b(1, p_{\underline{\xi}}^0(d))$ of a subject exposed to a dose $d \in [b_1, b_2]$. Our objective is then to choose

$E_{e_1}, \dots, E_{e_N} \in \mathcal{E}$ (possibly with repetitions) to provide independent observations $U_{e_1}, \dots, U_{e_N}$ for optimal inference on ξ_1 , and if the condition in Chernoff's theorem holds, then Part 2 of the theorem implies that k experiments in $\mathcal{E}$ will suffice for the large sample locally optimal design.

We will first replace Chernoff's condition by a set of more specific conditions and then discuss the interpretation of the optimality criterion.

Consider $E_d \in \mathcal{E}$, $U_d \sim b(1, p_\xi(d))$. The likelihood function $L(\xi) = p_\xi(d)^{U_d} (1 - p_\xi(d))^{1-U_d}$, hence

$$\ln L(\xi) = U_d \ln p_\xi(d) + (1 - U_d) \ln(1 - p_\xi(d)),$$

$$\frac{\partial \ln L(\xi)}{\partial \xi_j} = \frac{U_d - p_\xi(d)}{p_\xi(d)(1 - p_\xi(d))} \frac{\partial p_\xi(d)}{\partial \xi_j}, \quad j = 1, \dots, k$$

$$X(\xi) = \left[E_\xi \left(\frac{\partial \ln L(\xi)}{\partial \xi_i} \frac{\partial \ln L(\xi)}{\partial \xi_j} \right) \right]_{i,j=1,\dots,k}$$

$$= [x_{ij}(\xi)]_{i,j=1,\dots,k}$$

where

$$x_{ij}(\xi) = x_{ji}(\xi) = \frac{1}{p_\xi(d)(1 - p_\xi(d))} \frac{\partial p_\xi(d)}{\partial \xi_i} \frac{\partial p_\xi(d)}{\partial \xi_j}, \quad i, j = 1, \dots, k.$$

If ξ is the specified approximate value of ξ^0 , we will examine $X(\xi)$. Arranging the characteristic elements of $X(\xi)$ in an arbitrarily chosen order, we have the transpose of the $[k(k+1)]/2$ -vector which corresponds

to $X(\underline{\xi}) = (x_{11}(\underline{\xi}), x_{12}(\underline{\xi}), \dots, x_{1k}(\underline{\xi}), x_{22}(\underline{\xi}), \dots, x_{2k}(\underline{\xi}), \dots, x_{kk}(\underline{\xi}))$.

If $p_{\underline{\xi}}(d)$, $\partial p_{\underline{\xi}}(d)/\partial \xi_i$, $i = 1, \dots, k$ are continuous functions of $d \in [b_1, b_2]$, then R_1 is the continuous image of the compact set $[b_1, b_2]$ and is compact; therefore R , the convex hull of the closed and bounded R_1 , is closed and bounded and Chernoff's theorem applies.

The interpretation of the optimality criterion in Chernoff's theorem follows from Cramér's theorem [5]. The theorem can be stated for our case in the following way. Let $P = \{P_{\underline{\xi}}(d), 0 \leq d < \infty; \underline{\xi} \in Z\}$ be our model and $\underline{\xi}^0$ be the true value of the parameter corresponding to $p^0(d)$. Let the experiment of taking observations at k dose levels d_j , $j = 1, \dots, k$ in fixed proportions n_j/N , $j = 1, \dots, k$, be performed N times and the frequencies of response and non-response X_j , $n_j - X_j$, $j = 1, \dots, k$ be recorded, then, if

1. $\underline{\xi}^0$ is an interior point of D (a non-degenerate 'interval' of Z),
2. for any $\underline{\xi} \in D$,

$$p_{\underline{\xi}}(d_j) \in (c, 1-c), \quad c > 0, \quad j = 1, \dots, k,$$

$$\frac{\partial p_{\underline{\xi}}(d_i)}{\partial \xi_j}, \quad \frac{\partial^2 p_{\underline{\xi}}(d_i)}{\partial \xi_j \partial \xi_\ell} \quad \text{are all continuous in } \underline{\xi}, \quad i, j, \ell = 1, \dots, k,$$

and the matrix

$$\left[\frac{\partial p_{\underline{\xi}}(d_i)}{\partial \xi_j} \right]_{i,j=1,\dots,k} \quad \text{is of rank } k,$$

we have:

1. the likelihood equation has a unique solution $\hat{\underline{\zeta}}$ of $\underline{\zeta}$ such that

$$\hat{\underline{\zeta}} \xrightarrow[N \rightarrow \infty]{p} \underline{\zeta}^0$$

2. $\sqrt{N} (\hat{\underline{\zeta}} - \underline{\zeta}^0) \xrightarrow[N \rightarrow \infty]{d} \eta(0, X^{-1}(\underline{\zeta}^0))$ at all points of continuity of $X(\underline{\zeta}^0)$.

Hence the locally optimal design minimizes the variance of the asymptotic distribution of the maximum likelihood estimator of ζ_1 .

We will now assume the above conditions for the time being and derive the locally optimal design.

Our experiment has g groups of subjects with $n_i = \rho_i N$ subjects in group i exposed to a dose d_i , and

$$X_i = \sum_{j=1}^N U_{e_j} I_{[e_j=d_i]}, \quad i = 1, \dots, g.$$

The likelihood function

$$L(\underline{\zeta}) = \prod_{i=1}^g p_{\underline{\zeta}}(d_i)^{X_i} (1 - p_{\underline{\zeta}}(d_i))^{n_i - X_i},$$

hence

$$\ln L(\underline{\zeta}) = \sum_{i=1}^g \{X_i \ln p_{\underline{\zeta}}(d_i) + (n_i - X_i) \ln(1 - p_{\underline{\zeta}}(d_i))\}$$

$$\frac{\partial \ln L(\underline{\zeta})}{\partial \zeta_j} = \sum_{i=1}^g \left[\frac{X_i - n_i p_{\underline{\zeta}}(d_i)}{p_{\underline{\zeta}}(d_i)(1 - p_{\underline{\zeta}}(d_i))} \right] \frac{\partial p_{\underline{\zeta}}(d_i)}{\partial \zeta_j}, \quad j = 1, \dots, k.$$

Let $X(\underline{\zeta})$ be the Fisher information matrix of the mixed experiment which experiments at dose d_i in proportion ρ_i , $i = 1, \dots, g$.

... The Fisher information matrix of our experiment

$$\begin{aligned}
 &= N\mathbf{X}(\underline{\zeta}) \\
 &= \left[E_{\underline{\zeta}} \left(\frac{\partial \ln L(\underline{\zeta})}{\partial \zeta_i} \frac{\partial \ln L(\underline{\zeta})}{\partial \zeta_j} \right) \right]_{i,j=1,\dots,k} \\
 &= \left[\sum_{\ell=1}^g \frac{n_{\ell}}{p_{\underline{\zeta}}(d_{\ell})(1 - p_{\underline{\zeta}}(d_{\ell}))} \frac{\partial p_{\underline{\zeta}}(d_{\ell})}{\partial \zeta_i} \frac{\partial p_{\underline{\zeta}}(d_{\ell})}{\partial \zeta_j} \right]_{i,j=1,\dots,k} \\
 &= \mathbf{NBAB}'
 \end{aligned}$$

where

$$\mathbf{B} = \left[\frac{\partial p_{\underline{\zeta}}(d_j)}{\partial \zeta_i} \right]_{\substack{i=1,\dots,k \\ j=1,\dots,g}} \quad (2.2.1)$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & . & . & . \\ . & . & a_{\ell\ell} & . \\ . & . & . & . \\ . & . & . & 0 \\ 0 & . & . & 0 & a_{gg} \end{bmatrix}$$

with

$$a_{\ell\ell} = \frac{p_{\ell}}{p_{\underline{\zeta}}(d_{\ell})(1 - p_{\underline{\zeta}}(d_{\ell}))}, \quad \ell = 1, \dots, g.$$

If P is a one-parameter family, i.e., $k = 1$, the locally optimal design has $g = 1$ group of subjects. Hence

$$\rho_1 = 1$$

$$X(\underline{\xi}) = \frac{1}{p_{\underline{\xi}}(d_1)(1 - p_{\underline{\xi}}(d_1))} \left[\frac{\partial p_{\underline{\xi}}(d_1)}{\partial \xi_1} \right]^2$$

$$X^{11}(\underline{\xi}) = \frac{p_{\underline{\xi}}(d_1)(1 - p_{\underline{\xi}}(d_1))}{\frac{1}{2} \left[\frac{\partial p_{\underline{\xi}}(d_1)}{\partial \xi_1} \right]^2} \quad (2.2.2)$$

Therefore, all subjects are allocated at a dose $b_1 \leq d_1 \leq b_2$ which minimizes $X^{11}(\underline{\xi})$. Since the minimizing dose d_1 may depend on the parameter $\underline{\xi}$, we will denote it by $d_1(\underline{\xi})$, and we will denote the corresponding optimal allocation fraction by $\rho_1(\underline{\xi})$. Here of course $\rho_1(\underline{\xi}) = 1$.

Now if P is a k -parameter family $k \geq 2$, then g , the number of necessary experimental points is at most k and we seek the optimum doses and allocation fractions ρ_ℓ . Let B_{ij} be the minor of the (i, j) element of B , $\|B\|$ be the determinant of B , then

$$B^{-1} = \frac{1}{\|B\|} [(-1)^{i+j} B_{ji}]_{i,j=1,\dots,k}$$

$$X(\underline{\xi})^{-1} = B^{-1} A^{-1} B^{-1} = B^{-1} A^{-1} B^{-1}$$

$$= \left[\sum_{\ell=1}^k \frac{p_{\underline{\xi}}(d_\ell)(1 - p_{\underline{\xi}}(d_\ell))}{\rho_\ell} \frac{(-1)^{i+j} B_{i\ell} B_{j\ell}}{\|B\|^2} \right]_{i,j=1,\dots,k}$$

$$\begin{aligned}
x^{11}(\underline{\zeta}) &= \sum_{\ell=1}^k \frac{p_{\underline{\zeta}}(d_{\ell})(1-p_{\underline{\zeta}}(d_{\ell}))}{\rho_{\ell}} \frac{B_{1\ell}^2}{\|B\|^2} \\
&= \sum_{\ell=1}^k \frac{C_{\ell}^2}{\rho_{\ell}}
\end{aligned} \tag{2.2.3}$$

where

$$C_{\ell} = \sqrt{p_{\underline{\zeta}}(d_{\ell})(1-p_{\underline{\zeta}}(d_{\ell}))} \left| \frac{B_{1\ell}}{\|B\|} \right|, \quad \ell = 1, \dots, k \tag{2.2.4}$$

To find the optimal allocation of sample sizes, we minimize $x^{11}(\underline{\zeta})$ subject to $\sum_{\ell=1}^k \rho_{\ell} = 1$, i.e. with Lagrange multiplier λ , we minimize the expression

$$\sum_{\ell=1}^k \frac{C_{\ell}^2}{\rho_{\ell}} + \lambda \left(\sum_{\ell=1}^k \rho_{\ell} - 1 \right).$$

Setting the partial derivatives with respect to λ and ρ_{ℓ} , $\ell = 1, \dots, k$ to zeros, we have

$$\begin{aligned}
\sum_{\ell=1}^k \rho_{\ell} - 1 &= 0 \\
-\frac{C_{\ell}^2}{\rho_{\ell}^2} + \lambda &= 0, \quad \ell = 1, \dots, k.
\end{aligned}$$

Hence

$$\begin{aligned}
\sum_{\ell=1}^k \rho_{\ell} &= 1 \quad \text{and} \quad \rho_{\ell} = \frac{C_{\ell}}{\sqrt{\lambda}} \\
\Rightarrow \lambda &= \left(\sum_{\ell=1}^k C_{\ell} \right)^2 \quad \text{and} \quad \rho_{\ell} = \frac{C_{\ell}}{\sum_{\ell=1}^k C_{\ell}}, \quad \ell = 1, \dots, k
\end{aligned}$$

and with the optimal sample allocation,

$$x^{11}(\underline{\zeta}) = \left(\sum_{\ell=1}^k c_{\ell} \right)^2. \quad (2.2.5)$$

Hence if the approximate value of the parameter is $\underline{\zeta}$, the locally optimal design has $g = k$, doses at $d_i = d_i(\underline{\zeta})$, $b_1 \leq d_i(\underline{\zeta}) \leq b_2$, $i = 1, \dots, k$ which minimize $x^{11}(\underline{\zeta})$ or equivalently minimize

$$\sum_{\ell=1}^k \sqrt{p_{\underline{\zeta}}(d_{\ell})(1 - p_{\underline{\zeta}}(d_{\ell}))} \left| \frac{B_{1\ell}}{\|B\|} \right|,$$

and sample sizes allocated according to $\rho_{\ell} = \rho_{\ell}(\underline{\zeta})$ where

$$\rho_{\ell}(\underline{\zeta}) = \frac{\sqrt{p_{\underline{\zeta}}(d_{\ell})(1 - p_{\underline{\zeta}}(d_{\ell}))} \left| \frac{B_{1\ell}}{\|B\|} \right|}{\sum_{\ell=1}^k \sqrt{p_{\underline{\zeta}}(d_{\ell})(1 - p_{\underline{\zeta}}(d_{\ell}))} \left| \frac{B_{1\ell}}{\|B\|} \right|} \quad \begin{matrix} \ell = 1, \dots, k. \\ d_{\ell} = d_{\ell}(\underline{\zeta}) \\ \ell = 1, \dots, k \end{matrix} \quad (2.2.6)$$

The resulting design may depend on the model used, the parametrization of the model, and the parameter to be estimated. The local optimality of such design relies on the correctness of the model assumed. We will apply the above results to certain models in the next two sections with the assumption of the correctness of the model in each case and the sole intention of estimating either the low dose response rate or the safe dose.

2.3. Design for Estimating Low Dose Response Rate

Given a fixed low dose d_0 , say $0 < d_0 < b_1$, the problem in this section is to find the locally optimal design for estimating the low dose response rate $p^0(d_0)$. Two cases of the model proposed in Section 2.1 will be examined:

Case 1. with θ unknown and $0 < \beta < \infty$ known,

Case 2. with both θ and β unknown;

and the results, of course, depend on the correctness of these model assumptions.

Case 1. Assume $p^0(d) \in P_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty\}$,
 $0 < \beta < \infty$ known.

Let θ^0 be the true value of the parameter corresponding to $p^0(d)$.

1. Reparametrization.

Given $0 < d_0 < b_1$, let ξ be the parameter for $p^0(d_0)$. We can rewrite the model in terms of $0 < \xi < 1$,

$$\begin{aligned} \xi = 1 - e^{-\theta d_0^\beta} &\implies \theta = -\frac{\ln(1-\xi)}{d_0^\beta} \\ \implies p_\theta(d) &= 1 - e^{-\theta d^\beta} \\ &= 1 - (1-\xi)^{(d/d_0)^\beta} = p_\xi(d) \quad (2.3.1) \end{aligned}$$

Hence

$$P_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty\}$$

$$= \{p_\zeta(d) = 1 - (1 - \zeta)^{(d/d_0)^\beta}, 0 \leq d < \infty; 0 < \zeta < 1\}.$$

A function $p(d)$ in P_1 can be specified by a given θ value or a given ζ value depending on the parametrization used. The true dose response function $p^0(d)$ corresponds to $p_{\theta^0}(d)$ and $p_{\zeta^0}(d)$, and $\zeta^0 = p^0(d_0)$ is the true value of the parameter ζ .

2. Estimation and Design

For convenience, we will expand P_1 to $\bar{P}_1$ so that each point in the sample space will always correspond to a dose response function in our model.

$$\bar{P}_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 \leq \theta \leq \infty\}$$

$$= \{p_\zeta(d) = 1 - (1 - \zeta)^{(d/d_0)^\beta}, 0 \leq d < \infty; 0 \leq \zeta \leq 1\}.$$

From the expression for $p_\zeta(d)$,

$$\frac{\partial p_\zeta(d)}{\partial \zeta} = \left(\frac{d}{d_0}\right)^\beta (1 - \zeta)^{(d/d_0)^\beta - 1},$$

both $p_\zeta(d)$, $\partial p_\zeta(d)/\partial \zeta$ are continuous functions of $d \in [b_1, b_2]$, hence Chernoff's theorem applies, the locally optimal design for estimating ζ is a one point design.

We will assume an experiment with N subjects at dose $b_1 \leq d_1 \leq b_2$, and X_1 is the frequency of response $\sim b(N, p(d_1))$. Since the maximum likelihood estimator of $p(d_1)$ is X_1/N , and the parameter of interest ζ can be expressed in terms of $p(d_1)$:

$$p(d_1) = p_\zeta(d_1) = 1 - (1 - \zeta)^{(d_1/d_0)^\beta}$$

$$\Rightarrow \zeta = 1 - (1 - p(d_1))^{(d_0/d_1)^\beta},$$

the maximum likelihood estimator of ζ is:

$$\hat{\zeta} = 1 - \left(1 - \frac{X_1}{N}\right)^{(d_0/d_1)^\beta}$$

With $\zeta^0 = p^0(d_0)$ an interior point of the parameter space, and $0 < d_1 < \infty$, the conditions in Cramér's theorem hold. $x^{11}(\zeta)/N$ can be interpreted as the variance of the asymptotic distribution of $\hat{\zeta}$, and from (2.2.2),

$$x^{11}(\zeta) = \frac{p_\zeta(d_1)(1 - p_\zeta(d_1))}{\left[\left(\frac{d_1}{d_0}\right)^\beta (1 - \zeta)^{(d_1/d_0)^\beta - 1} \right]^2}.$$

Recall that d_0 is the inference dose, d_1 is the experimental dose and N , the sample size is large. Suppose $p^0(d_0) = c$, a given number, and the parameter ζ is known to be close to c , then

$$\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=c} = \sqrt{\frac{1}{N} \cdot \frac{p(\xi)(d_1)(1 - p_{\xi}(d_1))|_{\xi=c}}{\left[\left(\frac{d_1}{d_0} \right)^{\beta} (1-c) (d_1/d_0)^{\beta-1} \right]^2}}$$

is the asymptotic standard derivation of $\hat{\xi}$ based on a sample of size N . The value of d_1 which minimizes

$$\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=c},$$

i.e. $d_1(\xi)|_{\xi=c}$, depends on c ; but since $p^0(d_1) = p_{\xi}(d_1)|_{\xi=c}$,

$$\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=c} = \sqrt{\frac{1}{N} \left[(1-c) \ln(1-c) \right]^2 \frac{p^0(d_1)}{(1 - p^0(d_1)) [\ln(1 - p^0(d_1))]^2}}$$

which is minimized if

$$\frac{p^0(d_1)}{(1 - p^0(d_1)) [\ln(1 - p^0(d_1))]^2}$$

is minimized. Since the latter expression does not depend on c ,

$p^0(d_1(\xi))|_{\xi=c}$ is the same for all $0 < c < 1$.

To find the response rate of the optimizing experimental dose, we need examine graphs of

$$\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=c}$$

versus $p^0(d_1)$. Figure 1 shows how

$$\sqrt{\frac{x^{11}(\zeta)}{N}} \Big|_{\zeta=c}$$

varies with $p^0(d_1)$ for the following four cases (assuming the asymptotic expression holds even for small sample sizes like $N = 10, 30$):

	d_0	c	N
(A)	10^{-6}	10^{-5}	10
(B)	10^{-6}	10^{-5}	30
(C)	10^{-6}	10^{-8}	10
(D)	10^{-6}	10^{-8}	30

The cases for different sample sizes are mainly plotted for comparison with the results of the next chapter.

The graphs show the following results:

- (1) $p^0(d_1(\zeta)) \Big|_{\zeta=c} = .80$ for all $0 < c < 1$. Hence the optimal response rate to experiment at is .80 and it does not depend on where the inference dose d_0 is, or which function in P_1 is the true dose response function.
- (2) Since

$$\sqrt{\frac{x^{11}(\zeta)}{N}} \Big|_{\zeta=c}$$

does not depend on β , the optimal value of .80 is the same for all $0 < \beta < \infty$ known, in particular, for all the hit models.

(3) d_1 need not be located exactly at the level where $p^0(d_1) = .80$ since the graphs are very flat over the entire region for $p^0(d_1)$ from .10 to .95.

(4) The minimum value of

$$\frac{p^0(d_1)}{(1 - p^0(d_1))[\ln(1 - p^0(d_1))]^2}$$

is 1.54423, and the minimized value of

$$\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=c} \doteq c \sqrt{\frac{1.54423}{N}} \quad \text{for small } c.$$

Hence if ξ^0 is small and experimentation is at or near the optimum, $N \geq 1.54423$ would be adequate for estimating ξ in the sense that the standard error is not greater in order of magnitude than the parameter value provided that the model assumed is correct and the asymptotic expression holds well for the sample size.

Case 2. Assume $p^0(d) \in P_2 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty, 0 < \beta < \infty\}$.

Let

$$\underline{\theta}^0 = \begin{pmatrix} \theta^0 \\ \beta^0 \end{pmatrix}$$

be the true value of the parameter

$$\underline{\theta} = \begin{pmatrix} \theta \\ \beta \end{pmatrix}$$

corresponding to $p^0(d)$.

FIGURE 1. Asymptotic standard deviation of $\hat{\zeta}$ for different one point designs

$$\sqrt{x^{11}(\zeta)/N} \Big|_{\zeta=c} \text{ versus } p^0(d_1)$$

Case (A) $d_0 = 10^{-6}$, $p^0(d_0) = c = 10^{-5}$, $N = 10$

Case (B) $d_0 = 10^{-6}$, $p^0(d_0) = c = 10^{-5}$, $N = 30$

Case (C) $d_0 = 10^{-6}$, $p^0(d_0) = c = 10^{-8}$, $N = 10$

Case (D) $d_0 = 10^{-6}$, $p^0(d_0) = c = 10^{-8}$, $N = 30$

1. Reparametrization

Given $0 < d_0 < b_1$, let d_p be another chosen dose $d_0 < d_p < \infty$.

Let

$$\underline{\zeta} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

be the parameters for

$$\begin{pmatrix} p^0(d_0) \\ p^0(d_p) \end{pmatrix}.$$

We can rewrite the model in terms of ζ_1, ζ_2 :

$$\zeta_1 = 1 - e^{-\theta d_0^\beta}, \quad \zeta_2 = 1 - e^{-\theta d_p^\beta}, \quad \text{and} \quad 0 < \zeta_1 < \zeta_2 < 1.$$

Hence

$$\ln \theta + \beta \ln d_0 = \ln \ln(1 - \zeta_1)^{-1}$$

$$\ln \theta + \beta \ln d_p = \ln \ln(1 - \zeta_2)^{-1}$$

$$\Rightarrow \begin{pmatrix} \ln \theta \\ \beta \end{pmatrix} = \frac{1}{\ln d_p - \ln d_0} \begin{bmatrix} \ln d_p & -\ln d_0 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} \ln \ln(1 - \zeta_1)^{-1} \\ \ln \ln(1 - \zeta_2)^{-1} \end{pmatrix}$$

$$\Rightarrow \underline{p}_\theta(d) = 1 - \exp(-\exp[\ln \theta + \beta \ln d])$$

$$= 1 - \exp \left(-\exp \left[\frac{\ln d_p - \ln d}{\ln d_p - \ln d_0} \ln \ln \frac{1}{1 - \zeta_1} + \frac{\ln d - \ln d_0}{\ln d_p - \ln d_0} \ln \ln \frac{1}{1 - \zeta_2} \right] \right)$$

$$= \underline{p}_\zeta(d)$$

(2.3.2)

Hence

$$P_2 = \{ p_{\underline{\theta}}(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty, 0 < \beta < \infty \}$$

$$= \left\{ p_{\underline{\xi}}(d) = 1 - \exp \left(-\exp \left[\frac{\ln d_p - \ln d}{\ln d_p - \ln d_0} \ln \ln \frac{1}{1-\xi_1} + \frac{\ln d - \ln d_0}{\ln d_p - \ln d_0} \ln \ln \frac{1}{1-\xi_2} \right] \right), \right.$$

$$\left. 0 \leq d < \infty; 0 < \xi_1 < \xi_2 < 1 \right\}.$$

A function $p(d)$ in P_2 can be specified by

$$\underline{\theta} = \begin{pmatrix} \theta \\ \beta \end{pmatrix} \quad \text{or} \quad \underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}.$$

$p^0(d)$ corresponds to $p_{\underline{\theta}^0}(d)$ and $p_{\underline{\xi}^0}(d)$, and

$$\underline{\xi}^0 = \begin{pmatrix} \xi_1^0 \\ \xi_2^0 \end{pmatrix} = \begin{pmatrix} p^0(d_0) \\ p^0(d_p) \end{pmatrix}$$

is the true value of the parameter $\underline{\xi}$ with $0 < \xi_1^0 < \xi_2^0 < 1$.

2. Estimation and Design

Let

$$y = \ln \ln \frac{1}{1 - p_{\underline{\xi}}(d)}. \quad (2.3.3)$$

From (2.3.2),

$$\frac{\partial p_{\underline{\zeta}}(d)}{\partial \zeta_1} = e^{-e^y} e^y \left(\frac{\ln d_p - \ln d}{\ln d_p - \ln d_0} \right) \frac{1}{(1-\zeta_1) \ln \frac{1}{1-\zeta_1}} \quad (2.3.4)$$

$$\frac{\partial p_{\underline{\zeta}}(d)}{\partial \zeta_2} = e^{-e^y} e^y \left(\frac{\ln d - \ln d_0}{\ln d_p - \ln d_0} \right) \frac{1}{(1-\zeta_2) \ln \frac{1}{1-\zeta_2}} \quad (2.3.5)$$

$p_{\underline{\zeta}}(d)$, $\partial p_{\underline{\zeta}}(d)/\partial \zeta_1$, $\partial p_{\underline{\zeta}}(d)/\partial \zeta_2$ are all continuous functions of $d \in [b_1, b_2]$, hence Chernoff's theorem applies, the locally optimal design for estimating ζ_1 is a two point design.

Assume we have an experiment with $n_i = \rho_i N$ subjects at dose d_i , $i = 1, 2$, $0 < b_1 \leq d_1 < d_2 \leq b_2 < \infty$, and $X_i \sim b(n_i, p(d_i))$, $i = 1, 2$ are the frequencies of response for the two groups. Since $0 < d_0 < d_p < \infty$, $d_1 < d_2$, and $\underline{\zeta}^0$ is an interior point of the parameter space, the conditions in Crámer's theorem hold, hence the likelihood equation has a unique solution.

$$\begin{pmatrix} \hat{\zeta}_1 \\ \hat{\zeta}_2 \end{pmatrix} \xrightarrow[N \rightarrow \infty]{p} \begin{pmatrix} \zeta_1^0 \\ \zeta_2^0 \end{pmatrix}, \quad \text{and} \quad \sqrt{\frac{x^{11}(\underline{\zeta})}{N}} \bigg|_{\underline{\zeta}=\underline{\zeta}^0}$$

is the asymptotic standard deviation of $\hat{\zeta}_1$.

For finite sample sizes, the maximum likelihood estimator however is not defined over the whole sample space. By the same argument as in reparametrization,

$$p(d_1) = p_{\theta}(d_1) = 1 - e^{-\theta d_1^{\beta}}, \quad p(d_2) = p_{\theta}(d_2) = 1 - e^{-\theta d_2^{\beta}},$$

$0 < p(d_1) < p(d_2) < 1$. Equation (2.3.2) holds with d_0 replaced by d_1 , d_p replaced by d_2 , ξ_1 replaced by $p(d_1)$, and ξ_2 replaced by $p(d_2)$:

$$p_{\theta}(d) = 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1-p(d_1)} + \frac{\ln d - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1-p(d_2)} \right] \right)$$

$$0 \leq d < \infty, \quad 0 < p(d_1) < p(d_2) < 1. \quad (2.3.6)$$

Hence

$$\xi_1 = p_{\theta}(d_0) = 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1-p(d_1)} + \frac{\ln d_0 - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1-p(d_2)} \right] \right),$$

$$0 < p(d_1) < p(d_2) < 1.$$

If the frequencies of response $X_1 = x_1$, $X_2 = x_2$ are such that $0 < x_1/n_1 < x_2/n_2 < 1$, then the maximum likelihood estimate based on x_1, x_2 is:

$$\hat{\xi}_1_{x_1, x_2} = 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - \frac{x_1}{n_1}} + \frac{\ln d_0 - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - \frac{x_2}{n_2}} \right] \right).$$

Suppose, we define an estimator over the whole sample space,

$$\hat{\hat{\zeta}}_1 = \begin{cases} 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - \frac{X_1}{n_1}} + \frac{\ln d_0 - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - \frac{X_2}{n_2}} \right] \right), & \text{if } 0 < X_1/n_1 < X_2/n_2 < 1 \\ \frac{X_1 + X_2}{n_1 + n_2}, & \text{if } 0 \leq X_1/n_1 = X_2/n_2 \leq 1 \text{ or } 0 \leq X_2/n_2 < X_1/n_1 \leq 1 \\ 0, & \text{if } X_1/n_1 = 0, 0 < X_2/n_2 \leq 1 \text{ or } 0 < X_1/n_1 < 1, X_2/n_2 = 1 \end{cases}$$

then $\hat{\hat{\zeta}}_1 = \hat{\zeta}_1 + R$ where

$$R = \begin{cases} \frac{X_1 + X_2}{n_1 + n_2}, & \text{if } 0 \leq X_1/n_1 = X_2/n_2 \leq 1 \text{ or } 0 \leq X_2/n_2 < X_1/n_1 \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Since $0 \leq R \leq 1$, $n_i = \rho_i N$, $i = 1, 2$ with fixed ρ_i , $i = 1, 2$, and

$P(0 < X_1/n_1 < X_2/n_2 < 1) \xrightarrow[N \rightarrow \infty]{P} 1$, therefore $R \xrightarrow[N \rightarrow \infty]{P} 0$, $\hat{\hat{\zeta}}_1 = \hat{\zeta}_1 + o_p(1)$

and both $\hat{\zeta}_1$ and $\hat{\hat{\zeta}}_1$ have the same asymptotic distribution.

We will use $\hat{\xi}_1$ as our estimator for ξ_1 ; the optimality criterion in Chernoff's theorem can still be interpreted as the minimization of the asymptotic variance of our estimator. The following is the calculation for $x^{11}(\xi)$.

From (2.2.1), (2.3.3), (2.3.4), (2.3.5),

$$B = \begin{bmatrix} \frac{\partial p_{\xi}(d_1)}{\partial \xi_1} & \frac{\partial p_{\xi}(d_2)}{\partial \xi_1} \\ \frac{\partial p_{\xi}(d_1)}{\partial \xi_2} & \frac{\partial p_{\xi}(d_2)}{\partial \xi_2} \end{bmatrix}$$

$$B_{11} = \frac{\partial p_{\xi}(d_2)}{\partial \xi_2} = \left(\frac{\ln d_2 - \ln d_0}{\ln d_p - \ln d_0} \right) \frac{(1 - p_{\xi}(d_2)) \ln(1 - p_{\xi}(d_2))^{-1}}{(1 - \xi_2) \ln(1 - \xi_2)^{-1}}$$

$$B_{12} = \frac{\partial p_{\xi}(d_1)}{\partial \xi_2} = \left(\frac{\ln d_1 - \ln d_0}{\ln d_p - \ln d_0} \right) \frac{(1 - p_{\xi}(d_1)) \ln(1 - p_{\xi}(d_1))^{-1}}{(1 - \xi_2) \ln(1 - \xi_2)^{-1}}$$

$$\frac{\partial p_{\xi}(d_1)}{\partial \xi_1} = \left(\frac{\ln d_p - \ln d_1}{\ln d_p - \ln d_0} \right) \frac{(1 - p_{\xi}(d_1)) \ln(1 - p_{\xi}(d_1))^{-1}}{(1 - \xi_1) \ln(1 - \xi_1)^{-1}}$$

$$\frac{\partial p_{\xi}(d_2)}{\partial \xi_1} = \left(\frac{\ln d_p - \ln d_2}{\ln d_p - \ln d_0} \right) \frac{(1 - p_{\xi}(d_2)) \ln(1 - p_{\xi}(d_2))^{-1}}{(1 - \xi_1) \ln(1 - \xi_1)^{-1}}$$

$$\|B\| = \frac{(1 - p_{\xi}(d_1)) \ln(1 - p_{\xi}(d_1))^{-1} (1 - p_{\xi}(d_2)) \ln(1 - p_{\xi}(d_2))^{-1}}{(1 - \xi_1) \ln(1 - \xi_1)^{-1} (1 - \xi_2) \ln(1 - \xi_2)^{-1}} \cdot \frac{\ln d_2 - \ln d_1}{\ln d_p - \ln d_0}$$

From (2.2.3),

$$x^{11}(\xi) = [(1-\xi_1) \ln(1-\xi_1)^{-1}]^2$$

$$\left[\frac{1}{\rho_1} \left(\frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \right)^2 \frac{p_{\xi}(d_1)}{(1-p_{\xi}(d_1)) (\ln(1-p_{\xi}(d_1)))^{-1})^2} \right. \\ \left. + \frac{1}{\rho_2} \left(\frac{\ln d_1 - \ln d_0}{\ln d_2 - \ln d_1} \right)^2 \frac{p_{\xi}(d_2)}{(1-p_{\xi}(d_2)) (\ln(1-p_{\xi}(d_2)))^{-1})^2} \right],$$

and with optimal sample allocation, (2.2.4), (2.2.5) imply

$$x^{11}(\xi) = [(1-\xi_1) \ln(1-\xi_1)^{-1}]^2$$

$$\times \left[\sqrt{\frac{p_{\xi}(d_1)}{1-p_{\xi}(d_1)}} \frac{1}{\ln(1-p_{\xi}(d_1))^{-1}} \left| \frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \right| \right. \\ \left. + \sqrt{\frac{p_{\xi}(d_2)}{1-p_{\xi}(d_2)}} \frac{1}{\ln(1-p_{\xi}(d_2))^{-1}} \left| \frac{\ln d_1 - \ln d_0}{\ln d_2 - \ln d_1} \right| \right]^2$$

$$= [(1-\xi_1) \ln(1-\xi_1)^{-1}]^2$$

$$\times \left[\sqrt{\frac{p_{\xi}(d_1)}{1-p_{\xi}(d_1)}} \frac{1}{\ln(1-p_{\xi}(d_1))^{-1}} \left| \frac{\ln \ln(1-p_{\xi}(d_2))^{-1} - \ln \ln(1-\xi_1)^{-1}}{\ln \ln(1-p_{\xi}(d_2))^{-1} - \ln \ln(1-p_{\xi}(d_1))^{-1}} \right| \right. \\ \left. + \sqrt{\frac{p_{\xi}(d_2)}{1-p_{\xi}(d_2)}} \frac{1}{\ln(1-p_{\xi}(d_2))^{-1}} \left| \frac{\ln \ln(1-p_{\xi}(d_1))^{-1} - \ln \ln(1-\xi_1)^{-1}}{\ln \ln(1-p_{\xi}(d_2))^{-1} - \ln \ln(1-p_{\xi}(d_1))^{-1}} \right| \right]^2$$

Hence if

$$\underline{\zeta}^0 = \begin{pmatrix} \zeta_1^0 \\ \zeta_2^0 \end{pmatrix} = \begin{pmatrix} p^0(d_0) \\ p^0(d_p) \end{pmatrix} = \underline{c},$$

a given vector

$$\underline{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

and the parameter $\underline{\zeta}$ is known to be close to $\underline{c}$, then the asymptotic standard derivation of $\hat{\zeta}_1$ based on a sample of size N is:

$$\sqrt{\frac{x^{11}(\underline{\zeta})}{N}} \Big|_{\underline{\zeta}=\underline{c}}$$

$$= \frac{1}{N^{1/2}} [(1-c_1) \ln(1-c_1)^{-1}]$$

$$\left[\sqrt{\frac{p^0(d_1)}{1-p^0(d_1)}} \frac{1}{\ln(1-p^0(d_1))^{-1}} \left| \frac{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-c_1)^{-1}}{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-p^0(d_1))^{-1}} \right| \right]$$

$$\times \left[\sqrt{\frac{p^0(d_2)}{1-p^0(d_2)}} \frac{1}{\ln(1-p^0(d_2))^{-1}} \left| \frac{\ln \ln(1-p^0(d_1))^{-1} - \ln \ln(1-c_1)^{-1}}{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-p^0(d_1))^{-1}} \right| \right]$$

(2.3.7)

The values of d_1, d_2 which minimize $(\sqrt{x^{11}(\underline{\zeta})/N})|_{\underline{\zeta}=\underline{c}}$ depend on $\underline{c}$, but the response rates of the optimizing doses, $p^0(d_1(\underline{\zeta}))|_{\underline{\zeta}=\underline{c}}$ and $p^0(d_2(\underline{\zeta}))|_{\underline{\zeta}=\underline{c}}$, depend only on c_1 . Thus the optimal allocation fraction also depends on c_1 only:

from (2.2.6)

$$\rho_1(\underline{\zeta})|_{\underline{\zeta}=\underline{c}} = 1 - \rho_2(\underline{\zeta})|_{\underline{\zeta}=\underline{c}}$$

$$= \frac{1}{1 + \sqrt{\frac{p^0(d_2)(1-p^0(d_1))}{p^0(d_1)(1-p^0(d_2))} \frac{\ln \frac{1}{1-p^0(d_1)}}{\ln \frac{1}{1-p^0(d_2)}} \left| \frac{\ln \ln \frac{1}{1-p^0(d_1)} - \ln \ln \frac{1}{1-c_1}}{\ln \ln \frac{1}{1-p^0(d_2)} - \ln \ln \frac{1}{1-c_1}} \right|}}$$

$$\text{evaluated at } d_1 = d_1(\underline{\zeta})|_{\underline{\zeta}=\underline{c}}, d_2 = d_2(\underline{\zeta})|_{\underline{\zeta}=\underline{c}} \quad (2.3.8)$$

To find the locally optimal design, we will examine the true asymptotic standard deviation of $\hat{\zeta}_1$ for different designs. Since (2.3.7) does not depend on c_2 , whatever $\underline{\zeta}^0$ is, if ζ_1^0 has the same value c_1 , then for any given design, the asymptotic standard deviations of $\hat{\zeta}_1$ will be the same. Hence we need only investigate cases with different c_1 . We will explore two of the cases here:

$$(A) \quad d_0 = 10^{-6}, \quad c_1 = p^0(d_0) = 10^{-5}, \quad N = 30$$

$$(B) \quad d_0 = 10^{-6}, \quad c_1 = p^0(d_0) = 10^{-8}, \quad N = 30$$

Table 2 shows the results for Case (A) and Table 3, for Case (B).

For each of the tables, the two axes give the true response rates at the two experimental doses, i.e., $p^0(d_1)$, $p^0(d_2)$; the entry represents $\sqrt{x^{11}(\underline{\xi})/N} \big|_{\underline{\xi}=\underline{c}}$ for that particular design consisting of experimenting at $p^0(d_1)$, $p^0(d_2)$ with optimal allocation of sample sizes. In order to focus on the order of magnitude, an entry of $a_1 \cdot a_2 \times 10^{-b}$ is written as $-b \overset{a_1 a_2}{1}$, for example, the entry for the design $(p^0(d_1), p^0(d_2)) = (.1, .99)$ in Table 2 is 3.0×10^{-5} , written as $-5 \overset{30}{3}$. Designs in the observable region lie to the lower right of the dotted lines.

The tables give the following results:

- (1) for both cases, the optimal design in the observable region is at

$$d_1 = d_1(\underline{\xi}) \big|_{\underline{\xi}=\underline{c}}, \quad d_2 = d_2(\underline{\xi}) \big|_{\underline{\xi}=\underline{c}} \quad \text{where} \quad p^0(d_1(\underline{\xi})) \big|_{\underline{\xi}=\underline{c}} = .1, \\ p^0(d_2(\underline{\xi})) \big|_{\underline{\xi}=\underline{c}} = .99;$$

- (2) the surfaces of $(\sqrt{x^{11}(\underline{\xi})/N}) \big|_{\underline{\xi}=\underline{c}}$ are fairly flat near the optimum for both cases so that experimentation need not be performed at exactly the optimal values. Experimentation at $(p^0(d_1), p^0(d_2)) = (.1, .8), (.2, .9), \text{ or } (.3, .99)$ etc. are almost just as good;

- (3) from (2.3.8), the optimal allocation fraction $\rho_1(\underline{\xi}) \big|_{\underline{\xi}=\underline{c}}$ for Case (A) is .673, for Case (B) is .644 which are fairly close;

- (4) if the asymptotic expression approximates well and the model assumed is correct, the sample size required to estimate ξ_1^0 to within one order of magnitude in Case (A) is 263, for Case (B) is 674.

To examine how the optimal design changes with c_1 , we investigate (2.3.7), (2.3.8) for very small values of c_1 . If experimentation is to be performed in the observable region where $|\ln \ln(1-p^0(d_i))^{-1}|$, $i = 1, 2$ are negligible compared with $|\ln \ln(1-c_1)^{-1}|$, then

$$\left. \sqrt{\frac{x^{11}(\xi)}{N}} \right|_{\xi=c} = - \frac{1}{\sqrt{N}} c_1 \ln c_1 Q(p^0(d_1), p^0(d_2))$$

where

$$Q(p^0(d_1), p^0(d_2))$$

$$= \frac{1}{\left| \ln \ln \frac{1}{1-p^0(d_2)} - \ln \ln \frac{1}{1-p^0(d_1)} \right|} \times \left[\sqrt{\frac{p^0(d_1)}{1-p^0(d_1)}} \frac{1}{\ln(1-p^0(d_1))^{-1}} + \sqrt{\frac{p^0(d_2)}{1-p^0(d_2)}} \frac{1}{\ln(1-p^0(d_2))^{-1}} \right],$$

and

TABLE 2 $\sqrt{\frac{x^{11}(\xi)}{N}} \Big|_{\xi=\xi}$ for different two point designs ($p^0(d_1)$, $p^0(d_2)$)

Case (A) $d_0 = 10^{-6}$, $c_1 = p^0(d_0) = 10^{-5}$, $N = 30$

Upper Design Point $p^0(d_2)$

	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$	$1-10^{-8}$
$1-10^{-9}$	36	13	51	20	58	74	14	26	49	62	72	82	92	10	11	13	16	24	34	48	67	10	18	46
$1-10^{-8}$	36	13	46	17	58	41	55	95	18	23	26	30	34	38	43	49	59	94	14	20	31	55	14	
$1-10^{-7}$	35	12	45	16	58	29	26	37	68	85	10	11	13	14	16	19	23	38	58	93	17	45		
$1-10^{-6}$	35	12	44	16	58	24	15	16	27	33	39	47	50	57	65	75	93	16	27	52	14			
$1-10^{-5}$	35	12	44	16	58	23	11	85	12	14	16	19	21	24	27	32	41	77	15	43				
$1-10^{-4}$	35	12	43	16	58	22	97	57	59	68	78	88	99	11	13	15	20	45	13					
$1-10^{-3}$	34	12	43	16	58	22	92	46	37	40	45	50	56	63	74	90	12	40						
$1-10^{-2}$	34	12	43	16	58	22	91	43	30	31	33	37	41	48	57	73	12							
.9	33	12	42	15	58	22	93	45	30	32	36	41	49	62	87	16								
.8	33	12	42	15	58	23	95	46	33	36	42	52	68	98	19									
.7	32	11	41	15	58	23	97	48	37	42	52	69	10	20										
.6	32	11	41	15	58	23	98	50	41	50	68	10	21											
.5	32	11	41	15	58	23	10	52	47	64	98	20												
.4	31	11	41	15	58	23	10	56	57	89	19													
.3	31	11	40	15	58	23	11	60	75	16														
.2	30	11	40	15	58	24	11	68	13															
.1	29	10	39	15	58	24	12	89																
10^{-2}	25	91	35	14	58	28	21																	
10^{-3}	19	73	29	12	58	42																		
10^{-4}	12	47	20	10	58																			
10^{-5}	58	58	58	58																				
10^{-6}	22	12	94																					
10^{-7}	69	54																						
10^{-8}	25																							
	-1																							

Lower Design Point $p^0(d_1)$

TABLE 3

$$\sqrt{\frac{x^{11}(\xi)}{N}} \quad \xi = \xi$$

for different two point designs ($p^0(d_1)$, $p^0(d_2)$)Case (B) $d_0 = 10^{-6}$, $c_1 = p^0(d_0) = 10^{-8}$, $N = 30$ Lower Design Point $p^0(d_1)$ Upper Design Point $p^0(d_2)$

	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$	$1-10^{-8}$
$1-10^{-9}$	52 -5	18 -5	68 -6	31 -6	22 -6	24 -6	34 -6	51 -6	86 -6	10 -5	12 -5	13 -5	15 -5	16 -5	18 -5	21 -5	24 -5	37 -5	52 -5	71 -5	10 -4	15 -4	27 -4	69 -4
$1-10^{-8}$	52 -5	18 -5	66 -6	26 -6	13 -6	11 -6	13 -6	19 -6	31 -6	38 -6	44 -6	49 -6	55 -6	61 -6	68 -6	77 -6	92 -6	14 -5	21 -5	30 -5	46 -5	82 -5	21 -4	
$1-10^{-7}$	52 -5	18 -5	65 -6	24 -6	10 -6	60 -7	55 -7	72 -7	12 -6	14 -6	17 -6	19 -6	21 -6	23 -6	26 -6	30 -6	36 -6	58 -6	88 -6	14 -5	25 -5	67 -5		
$1-10^{-6}$	52 -5	18 -5	65 -6	24 -6	92 -7	43 -7	29 -7	31 -7	46 -7	56 -7	65 -7	73 -7	82 -7	91 -7	10 -6	12 -6	14 -6	25 -6	41 -6	78 -6	21 -5			
$1-10^{-5}$	52 -5	18 -5	65 -6	24 -6	89 -7	37 -7	19 -7	15 -7	20 -7	24 -7	27 -7	31 -7	34 -7	38 -7	43 -7	51 -7	63 -7	12 -6	23 -6	65 -6				
$1-10^{-4}$	52 -5	18 -5	65 -6	24 -6	88 -7	35 -7	16 -7	95 -8	98 -8	11 -7	13 -7	14 -7	16 -7	18 -7	20 -7	24 -7	31 -7	69 -7	20 -6					
$1-10^{-3}$	52 -5	18 -5	65 -6	24 -6	88 -7	34 -7	14 -7	75 -8	60 -8	65 -8	72 -8	80 -8	89 -8	10 -7	12 -7	14 -7	19 -7	61 -7	-----					
$1-10^{-2}$	52 -5	18 -5	65 -6	24 -6	89 -7	34 -7	14 -7	68 -8	47 -8	49 -8	53 -8	58 -8	65 -8	75 -8	89 -8	11 -7	18 -7							
.9	52 -5	18 -5	66 -6	24 -6	90 -7	35 -7	15 -7	71 -8	49 -8	51 -8	57 -8	65 -8	78 -8	98 -8	14 -7	25 -7								
.8	51 -5	18 -5	66 -6	24 -6	91 -7	36 -7	15 -7	74 -8	53 -8	59 -8	68 -8	83 -8	11 -7	16 -7	30 -7									
.7	51 -5	18 -5	66 -6	24 -6	92 -7	36 -7	16 -7	78 -8	59 -8	68 -8	84 -8	11 -7	16 -7	32 -7										
.6	51 -5	18 -5	66 -6	24 -6	93 -7	37 -7	16 -7	82 -8	67 -8	82 -8	11 -7	17 -7	33 -7											
.5	51 -5	18 -5	66 -6	25 -6	94 -7	38 -7	16 -7	86 -8	78 -8	10 -7	16 -7	33 -7												
.4	51 -5	18 -5	66 -6	25 -6	95 -7	38 -7	17 -7	93 -8	95 -8	15 -7	31 -7													
.3	51 -5	18 -5	67 -6	25 -6	96 -7	39 -7	18 -7	10 -7	13 -7	27 -7														
.2	51 -5	18 -5	67 -6	25 -6	98 -7	41 -7	19 -7	12 -7	22 -7															
.1	51 -5	18 -5	67 -6	26 -6	10 -7	43 -7	21 -7	16 -7																
10^{-2}	49 -5	18 -5	69 -6	27 -6	12 -6	58 -7	44 -7																	
10^{-3}	48 -5	18 -5	72 -6	31 -6	15 -6	11 -6																		
10^{-4}	46 -5	18 -5	78 -6	38 -6	29 -6																			
10^{-5}	43 -5	18 -5	89 -6	66 -6																				
10^{-6}	39 -5	18 -5	13 -5																					
10^{-7}	32 -5	18 -5																						
10^{-8}	18 -5																							

$$\rho_1(\underline{\xi}) \Big|_{\underline{\xi}=\underline{c}}$$

$$= \frac{1}{1 + \sqrt{\frac{p^0(d_2)(1-p^0(d_1))}{p^0(d_1)(1-p^0(d_2))} \frac{\ln(1-p^0(d_1))^{-1}}{\ln(1-p^0(d_2))^{-1}}}} \Bigg|_{\substack{d_1=d_1(\underline{\xi})|_{\underline{\xi}=\underline{c}} \\ d_2=d_2(\underline{\xi})|_{\underline{\xi}=\underline{c}}}} \quad (2.3.9)$$

Table 4 shows the values of $Q(p^0(d_1), p^0(d_2))$ for different designs $(p^0(d_1), p^0(d_2))$. The following results are observed:

- (1) the optimal design in the observable region is at $d_1 = d_1(\underline{\xi})|_{\underline{\xi}=\underline{c}}$
 $d_2 = d_2(\underline{\xi})|_{\underline{\xi}=\underline{c}}$ where $(p^0(d_1(\underline{\xi})), p^0(d_2(\underline{\xi})))|_{\underline{\xi}=\underline{c}} = (.1, .99)$;
- (2) the surface is fairly flat near the optimum so that experimentation can be performed at other points in the vicinity;
- (3) from (2.3.9), the optimal allocation fraction $\rho_1(\underline{\xi})|_{\underline{\xi}=\underline{c}}$ for experimenting at the optimal design is approximately .594;
- (4) $Q(p^0(d_1(\underline{\xi})), p^0(d_2(\underline{\xi})))|_{\underline{\xi}=\underline{c}} \doteq 1.4095$ at the optimal design, therefore, if the asymptotic expression holds well, in order to estimate ξ_1^0 to the right order of magnitude: $-1.4095c_1 \ln c_1 / \sqrt{N} < c_1$
 $\Rightarrow N > (-1.4095 \ln c_1)^2$. Hence under this model assumption, a few hundred samples would be necessary if c_1 is in the range from about 10^{-3} to 10^{-10} , a few thousand samples would be necessary if c_1 is in the range from about 10^{-10} to 10^{-31} .

TABLE 4 $Q(p^0(d_1), p^0(d_2))$ for different two point designs $(p^0(d_1), p^0(d_2))$

Lower Design Point $p^0(d_1)$ →

	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$	$1-10^{-8}$
$1-10^{-9}$	14	54	24	15	13	13	16	20	29	34	38	41	45	49	54	60	70	10	14	19	26	39	69	18
$1-10^{-8}$	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	4
$1-10^{-7}$	14	49	19	92	60	53	58	74	11	12	14	15	17	18	20	22	26	39	56	80	12	21	55	
$1-10^{-6}$	3	2	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	
$1-10^{-5}$	14	48	18	72	36	25	24	28	40	46	52	57	63	69	76	86	10	16	24	37	66	17		
$1-10^{-4}$	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	3	
$1-10^{-3}$	14	48	17	65	27	15	11	11	15	18	20	22	25	27	30	34	41	68	11	21	55			
$1-10^{-2}$	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2			
.9	14	48	17	63	25	11	63	53	65	75	84	93	10	11	13	15	18	32	63	17				
.8	3	2	2	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2			
.7	14	48	17	63	24	97	47	31	31	35	39	43	48	53	60	69	88	19	54					
.6	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1				
.5	14	49	18	64	24	94	41	22	19	20	22	24	26	29	33	40	54	17						
.4	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1						
.3	14	50	18	65	24	95	40	20	14	15	16	17	19	22	26	32	50							
.2	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0							
.1	15	52	19	68	26	10	43	21	14	15	17	19	23	29	40	71								
10^{-2}	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0								
10^{-3}	15	53	19	70	26	10	45	22	16	18	20	25	32	46	87									
10^{-4}	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0									
10^{-5}	15	54	19	72	27	11	46	24	18	21	26	33	49	95										
10^{-6}	3	2	2	1	1	0	0	0	0	0	0	0	0	0										
10^{-7}	15	55	20	73	28	11	48	25	21	25	34	50	10											
10^{-8}	3	2	2	1	1	0	0	0	0	0	0	0	1											
10^{-9}	16	55	20	74	28	11	51	27	24	32	49	10												
10^{-10}	3	2	2	1	1	0	0	0	0	0	0	0												
10^{-11}	16	56	20	76	29	12	53	30	30	46	96													
10^{-12}	3	2	2	1	1	0	0	0	0	0	0													
10^{-13}	16	58	21	78	30	12	57	33	41	87														
10^{-14}	3	2	2	1	1	0	0	0	0	0														
10^{-15}	16	59	22	81	32	13	63	39	72															
10^{-16}	3	2	2	1	1	0	0	0	0															
10^{-17}	17	62	23	87	34	15	75	56																
10^{-18}	3	2	2	1	1	0	0																	
10^{-19}	20	72	28	11	47	24	18																	
10^{-20}	3	2	2	2	1																			
10^{-21}	23	87	35	15	76	57																		
10^{-22}	3	2	2	2	1																			
10^{-23}	28	11	47	24	18																			
10^{-24}	3	3	2	2	2																			
10^{-25}	35	15	76	57																				
10^{-26}	3	3	2	2																				
10^{-27}	47	24	18																					
10^{-28}	3	3																						
10^{-29}	76	57																						
10^{-30}	3																							
10^{-31}	18																							
10^{-32}	4																							

↑ Upper Design Point $p^0(d_2)$

These results thus show that with the extreme value distribution model in which both parameters are unknown, the locally optimal design for inference on the response rate at a low dose (say, $p^0(d_0) < 10^{-3}$) is relatively stable. It involves experimenting at two dose levels, $p^0(d_1) = .1$ and $p^0(d_2) = .99$. Approximately 60-70 percent of the total sample size should be allocated to the lower dose, and for $10^{-31} < p^0(d_0) < 10^{-3}$, hundreds to thousands of animals (depending on $p^0(d_0)$) need be used to estimate the low dose response rate to the right order of magnitude.

2.4. Design for Estimating Safe Dose

Given a fixed low response rate p_0 , say $0 < p_0 < 10^{-3}$, the problem in this section is to find the locally optimal design for estimating the safe dose $d^0(p_0)$ defined by $p^0(d^0(p_0)) = p_0$. Two cases of the model proposed in Section 2.1 will be examined.

Case 1. with θ unknown and $0 < \beta < \infty$ known,

Case 2. with both θ and β unknown.

The results, again, depend on the correctness of these model assumptions.

Case 1. Assume $p^0(d) \in P_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty\}$,
 $0 < \beta < \infty$ known.

Let θ^0 be the value of the parameter corresponding to $p^0(d)$.

1. Reparametrization

We wish to estimate $d^0(p_0)$ where, say, $0 < p_0 < 10^{-3}$; let ξ be the parameter for $d^0(p_0)$, we can express the model in terms of $0 < \xi < \infty$,

$$p_0 = 1 - e^{-\theta \xi^\beta} \Rightarrow \theta = \frac{-\ln(1-p_0)}{\xi^\beta}$$

$$\Rightarrow p_\theta(d) = 1 - e^{-\theta d^\beta} = 1 - (1-p_0)^{(d/\xi)^\beta} = p_\xi(d) \quad (2.4.1)$$

$$\therefore P_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty\}$$

$$= \{p_\xi(d) = 1 - (1-p_0)^{(d/\xi)^\beta}, 0 \leq d < \infty; 0 < \xi < \infty\}.$$

A function $p(d)$ in P_1 can be specified by a given θ or ξ value depending on the parametrization. The true dose response function $p^0(d)$ corresponds to $p_{\theta^0}(d)$ and $p_{\xi^0}(d)$, and $\xi^0 = d^0(p_0)$ is the true value of the parameter ξ .

2. Estimation and Design

Again for convenience, we will expand P_1 to $\bar{P}_1$ so that each point in the sample space will always correspond to a dose response function in our model.

$$\bar{P}_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 \leq \theta \leq \infty\}$$

$$= \{p_\xi(d) = 1 - (1-p_0)^{(d/\xi)^\beta}, 0 \leq d < \infty; 0 \leq \beta \leq \infty\}$$

From the expression for $p_\xi(d)$ in (2.4.1),

$$\frac{\partial p_\xi(d)}{\partial \xi} = (1-p_0) (d/\xi)^\beta \ln(1-p_0) \frac{d^\beta}{\xi^{\beta+1}}.$$

Both $p_\xi(d)$, $\partial p_\xi(d)/\partial \xi$ are continuous functions of $d \in [b_1, b_2]$, hence Chernoff's theorem applies, the locally optimal design for estimating ξ is a one point design.

We will assume an experiment with N subjects at dose $b_1 \leq d_1 \leq b_2$, and $X_1 \sim b(N, p(d_1))$ is the frequency of response. Since the maximum likelihood estimator of $p(d_1)$ is X_1/N , and the parameter of interest ξ can be expressed in terms of $p(d_1)$, from (2.4.1) with $d = d_1$,

$$p(d_1) = p_\xi(d_1) = 1 - (1-p_0) (d_1/\xi)^\beta \implies \xi = d_1 \left[\frac{\ln(1-p_0)}{\ln(1-p(d_1))} \right]^{1/\beta}$$

hence the maximum likelihood estimator of ξ is:

$$\hat{\xi} = d_1 \left[\frac{\ln(1-p_0)}{\ln[1-(X_1/N)]} \right]^{1/\beta}.$$

With $\xi^0 = d^0(p_0)$ an interior point of the parameter space, and $0 < d_1 < \infty$, the conditions in Cramér's theorem hold. $x^{11}(\xi)/N$ can be interpreted as the variance of the asymptotic distribution of $\hat{\xi}$, and from (2.2.2),

$$x^{11}(\xi) = \frac{p_\xi(d_1)(1 - p_\xi(d_1))}{\left[(1-p_0) (d_1/\xi)^\beta \ln(1-p_0) \frac{d_1^\beta}{\xi^{\beta+1}} \right]^2}$$

Since d_1 is the experimental dose, N is large, and if the parameter ζ is known to be close to the true value $\zeta^0 = c$, then the asymptotic standard deviation of $\hat{\zeta}$ based on a sample of size N is:

$$\begin{aligned} \sqrt{\frac{11(\zeta)}{N}} \Big|_{\zeta=c} &= \sqrt{\left\{ \frac{1}{N} \frac{p_{\zeta}(d_1)(1-p_{\zeta}(d_1))}{[(1-p_{\zeta}(d_1)) \ln(1-p_{\zeta}(d_1))(\beta/\zeta)]^2} \right\} \Big|_{\zeta=c}} \\ &= \sqrt{\left\{ \frac{1}{N} \left(\frac{c}{\beta}\right)^2 \frac{p^0(d_1)}{(1-p^0(d_1))[\ln(1-p^0(d_1))]^2} \right\}} \end{aligned}$$

and is minimized if

$$\frac{p^0(d_1)}{(1-p^0(d_1)) [\ln(1-p^0(d_1))]^2}$$

is minimized. Hence even though the value of d_1 which minimizes $\sqrt{x^{11}(\zeta)/N} \Big|_{\zeta=c}$, i.e., $d_1(\zeta) \Big|_{\zeta=c}$ depends on c , the response rate at the optimal dose $p^0(d_1(\zeta)) \Big|_{\zeta=c}$ does not depend on c . Further the latter expression is the same as that for estimating the low dose response rate, thus similar results hold:

- (1) $p^0(d_1(\zeta)) \Big|_{\zeta=c} = .80$ for all $0 < c < \infty$. The optimum response rate to experiment at is .80 and it does not depend on where the safe rate p_0 is, or which function in P_1 is the true dose response function.
- (2) The optimum value of .80 is the same for all $0 < \beta < \infty$ known, i.e. for all the hit models.

- (3) $p^0(d_1)$ need not be exactly at .80 since not much is lost for experimenting at $p^0(d_1)$ from .1 to .95.
- (4) The minimum value of

$$\frac{p^0(d_1)}{(1-p^0(d_1))[\ln(1-p^0(d_1))]^2}$$

is 1.54423, therefore, $\sqrt{x^{11}(\zeta)/N} \Big|_{\zeta=c} = c/(\sqrt{N} \beta) \sqrt{1.54423}$

when the experiment is carried out at or near the optimum.

The sample size required to estimate ζ to the right order of magnitude depends on β : $N > 1.54423/\beta^2$. Hence if the model is really a hit model with known $\beta \geq 1$, a sample size $N > 1.54423$ is adequate for estimating ζ in the sense that the standard error is not greater in order of magnitude than the parameter value.

Case 2. Assume $p^0(d) \in P_2 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty, 0 < \beta < \infty\}$

Let

$$\underline{\theta}^0 = \begin{pmatrix} \theta^0 \\ \beta^0 \end{pmatrix}$$

be the true value of the parameter

$$\underline{\theta} = \begin{pmatrix} \theta \\ \beta \end{pmatrix}$$

corresponding to $p^0(d)$.

1. Reparametrization

Given $0 < p_0 < 10^{-3}$, let p_t be another chosen response rate with $p_0 < p_t < 1$.

Let

$$\underline{\zeta} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

be the parameters for

$$\begin{pmatrix} d^0(p_0) \\ d^0(p_t) \end{pmatrix},$$

$$\therefore p_0 = 1 - e^{-\theta \zeta_1^\beta}, p_t = 1 - e^{-\theta \zeta_2^\beta}.$$

Rewrite (2.3.2) with ζ_1 replaced by p_0 , ζ_2 replaced by p_t , d_0 replaced by ζ_1 and d_p replaced by ζ_2 ,

$$p_{\underline{\zeta}}(d) = 1 - \exp \left(- \exp \left[\frac{\ln \zeta_2 - \ln d}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_0} + \frac{\ln d - \ln \zeta_1}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_t} \right] \right). \quad (2.4.2)$$

Hence

$$P_2 = \{p_{\underline{\zeta}}(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty: 0 < \theta < \infty, 0 < \beta < \infty\}$$

$$= \left\{ p_{\underline{\zeta}}(d) = 1 - \exp \left(- \exp \left[\frac{\ln \zeta_2 - \ln d}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_0} + \frac{\ln d - \ln \zeta_1}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_t} \right] \right), 0 \leq d < \infty: 0 < \zeta_1 < \zeta_2 < \infty \right\}.$$

A function in P_2 can be specified by a

$$\underline{\theta} = \begin{pmatrix} \theta \\ \beta \end{pmatrix} \quad \text{or} \quad \underline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

value. $p^0(d)$ corresponds to $p_{\underline{\theta}^0}(d)$ and $p_{\underline{\xi}^0}(d)$, and

$$\underline{\xi}^0 = \begin{pmatrix} \xi_1^0 \\ \xi_2^0 \end{pmatrix} = \begin{pmatrix} d^0(p_0) \\ d^0(p_t) \end{pmatrix}$$

is the true value of the parameter $\underline{\xi}$ with $0 < \xi_1^0 < \xi_2^0 < \infty$.

2. Estimation and Design

Let

$$y = \ln \ln \frac{1}{1-p_{\underline{\xi}}(d)} \quad (2.4.3)$$

From (2.4.2),

$$\begin{aligned} \frac{\partial p_{\underline{\xi}}(d)}{\partial \xi_1} &= e^{-e^y} e^y \left(\ln \ln \frac{1}{1-p_t} - \ln \ln \frac{1}{1-p_0} \right) \\ &\quad \times \frac{(\ln d - \ln \xi_2)}{(\ln \xi_2 - \ln \xi_1)^2} \cdot \frac{1}{\xi_1} \end{aligned} \quad (2.4.4)$$

$$\begin{aligned} \frac{\partial p_{\underline{\xi}}(d)}{\partial \xi_2} &= e^{-e^y} e^y \left(\ln \ln \frac{1}{1-p_0} - \ln \ln \frac{1}{1-p_t} \right) \\ &\quad \times \frac{(\ln d - \ln \xi_1)}{(\ln \xi_2 - \ln \xi_1)^2} \cdot \frac{1}{\xi_2} \end{aligned} \quad (2.4.5)$$

$p_{\underline{\zeta}}(d)$, $\partial p_{\underline{\zeta}}(d)/\partial \zeta_1$, $\partial p_{\underline{\zeta}}(d)/\partial \zeta_2$ are all continuous functions of $d \in [b_1, b_2]$ hence Chernoff's theorem holds, the locally optimal design for estimating $\underline{\zeta}_1$ is a two point design.

Let the experiment have $n_i = \rho_i N$ subjects at dose d_i , $i = 1, 2$, $b_1 \leq d_1 < d_2 \leq b_2$, and $X_i \sim b(n_i, p(d_i))$, $i = 1, 2$ are the frequencies of response for the two groups. Since $\underline{\zeta}^0$ is an interior point of the parameter space and the other conditions in Cramér's theorem hold, the likelihood equation has a unique solution

$$\begin{pmatrix} \hat{\zeta}_1 \\ \hat{\zeta}_2 \end{pmatrix} \xrightarrow[N \rightarrow \infty]{P} \begin{pmatrix} \zeta_1^0 \\ \zeta_2^0 \end{pmatrix}, \quad \text{and} \quad \sqrt{\frac{x^{11}(\underline{\zeta})}{N}} \bigg|_{\underline{\zeta} = \underline{\zeta}^0}$$

is the asymptotic standard deviation of $\hat{\zeta}_1$ when $\underline{\zeta}^0$ obtains.

From (2.4.2)

$$\begin{aligned} \ln \ln \frac{1}{1-p(d_1)} &= \ln \ln \frac{1}{1-p_{\underline{\zeta}}(d_1)} \\ &= \frac{\ln \zeta_2 - \ln d_1}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_0} \\ &\quad + \frac{\ln d_1 - \ln \zeta_1}{\ln \zeta_2 - \ln \zeta_1} \ln \ln \frac{1}{1-p_t} \end{aligned}$$

$$\begin{aligned}
& \ln \ln \frac{1}{1-p(d_2)} \\
&= \ln \ln \frac{1}{1-p_{\xi}(d_2)} = \frac{\ln \xi_2 - \ln d_2}{\ln \xi_2 - \ln \xi_1} \ln \ln \frac{1}{1-p_0} + \frac{\ln d_2 - \ln \xi_1}{\ln \xi_2 - \ln \xi_1} \ln \ln \frac{1}{1-p_t} \\
&\Rightarrow \ln \xi_1 = \frac{\ln \ln(1-p(d_2))^{-1} - \ln \ln(1-p_0)^{-1}}{\ln \ln(1-p(d_2))^{-1} - \ln \ln(1-p(d_1))^{-1}} \ln d_1 \\
&\quad + \frac{\ln \ln(1-p_0)^{-1} - \ln \ln(1-p(d_1))^{-1}}{\ln \ln(1-p(d_2))^{-1} - \ln \ln(1-p(d_1))^{-1}} \ln d_2 .
\end{aligned}$$

If the frequencies of response $X_1 = x_1$, $X_2 = x_2$ are such that $0 < x_1/n_1 < x_2/n_2 < 1$, the maximum likelihood estimator based on x_1, x_2 is:

$$\begin{aligned}
\hat{\xi}_{1, x_1, x_2} &= d_1 \left[\frac{\ln \ln(1-(x_2/n_2))^{-1} - \ln \ln(1-p_0)^{-1}}{\ln \ln(1-(x_2/n_2))^{-1} - \ln \ln(1-(x_1/n_1))^{-1}} \right] \\
&\quad \times d_2 \left[\frac{\ln \ln(1-p_0)^{-1} - \ln \ln(1-(x_1/n_1))^{-1}}{\ln \ln(1-(x_2/n_2))^{-1} - \ln \ln(1-(x_1/n_1))^{-1}} \right]
\end{aligned}$$

For finite sample sizes, the maximum likelihood estimator is not defined over the whole sample space, but we can define an estimator over the whole sample space as:

$$\hat{\xi}_1 = \begin{cases} \hat{\xi}_1^{X_1, X_2} & , \quad \text{if } 0 < X_1/n_1 < X_2/n_2 < 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

then $\hat{\xi}_1$ and $\hat{\xi}_1$ will have the same asymptotic distribution and hence the same asymptotic variance.

The calculation for the asymptotic variance is as follows:

from (2.2.1), (2.4.3), (2.4.4), (2.4.5), (2.2.4), (2.2.5),

$$B = \begin{bmatrix} \frac{\partial p_{\xi}(d_1)}{\partial \xi_1} & \frac{\partial p_{\xi}(d_2)}{\partial \xi_1} \\ \frac{\partial p_{\xi}(d_1)}{\partial \xi_2} & \frac{\partial p_{\xi}(d_2)}{\partial \xi_2} \end{bmatrix}$$

$$B_{11} = \frac{\partial p_{\xi}(d_2)}{\partial \xi_2}$$

$$= (1-p_{\xi}(d_2)) \ln \frac{1}{(1-p_{\xi}(d_2))} \left(\ln \ln \frac{1}{1-p_0} - \ln \ln \frac{1}{1-p_t} \right) \\ \times \frac{(\ln d_2 - \ln \xi_1)}{(\ln \xi_2 - \ln \xi_1)^2} \frac{1}{\xi_2}$$

$$B_{12} = \frac{\partial p_{\xi}(d_1)}{\partial \xi_2}$$

$$= (1-p_{\xi}(d_1)) \ln \frac{1}{(1-p_{\xi}(d_1))} \left(\ln \ln \frac{1}{1-p_0} - \ln \ln \frac{1}{1-p_t} \right) \\ \times \frac{(\ln d_1 - \ln \xi_1)}{(\ln \xi_2 - \ln \xi_1)^2} \frac{1}{\xi_2}$$

$$\begin{aligned} \frac{\partial p_{\underline{\zeta}}(d_1)}{\partial \zeta_1} &= (1-p_{\underline{\zeta}}(d_1)) \ln \frac{1}{(1-p_{\underline{\zeta}}(d_1))} \left(\ln \ln \frac{1}{1-p_t} - \ln \ln \frac{1}{1-p_0} \right) \\ &\quad \times \frac{(\ln d_1 - \ln \zeta_2)}{(\ln \zeta_2 - \ln \zeta_1)^2} \frac{1}{\zeta_1} \end{aligned}$$

$$\begin{aligned} \frac{\partial p_{\underline{\zeta}}(d_2)}{\partial \zeta_1} &= (1-p_{\underline{\zeta}}(d_2)) \ln \frac{1}{(1-p_{\underline{\zeta}}(d_2))} \left(\ln \ln \frac{1}{1-p_t} - \ln \ln \frac{1}{1-p_0} \right) \\ &\quad \times \frac{(\ln d_2 - \ln \zeta_2)}{(\ln \zeta_2 - \ln \zeta_1)^2} \frac{1}{\zeta_1} \end{aligned}$$

$$\begin{aligned} \|B\| &= (1-p_{\underline{\zeta}}(d_1)) \ln \frac{1}{(1-p_{\underline{\zeta}}(d_1))} (1-p_{\underline{\zeta}}(d_2)) \\ &\quad \times \ln \frac{1}{(1-p_{\underline{\zeta}}(d_2))} \left(\ln \ln \frac{1}{1-p_t} - \ln \ln \frac{1}{1-p_0} \right)^2 \\ &\quad \times \frac{(\ln d_2 - \ln d_1)}{(\ln \zeta_2 - \ln \zeta_1)^3 \zeta_1 \zeta_2} \end{aligned}$$

therefore with optimal sample allocation,

$$\begin{aligned}
x^{11}(\underline{\zeta}) &= \left[\frac{\zeta_1 (\ln \zeta_2 - \ln \zeta_1)}{\ln \ln(1-p_t)^{-1} - \ln \ln(1-p_0)^{-1}} \right]^2 \\
&\times \left[\sqrt{\frac{p_{\underline{\zeta}}(d_1)}{1-p_{\underline{\zeta}}(d_1)}} \cdot \frac{1}{\ln(1-p_{\underline{\zeta}}(d_1))^{-1}} \left| \frac{\ln d_2 - \ln \zeta_1}{\ln d_2 - \ln d_1} \right| \right. \\
&\quad \left. + \sqrt{\frac{p_{\underline{\zeta}}(d_2)}{1-p_{\underline{\zeta}}(d_2)}} \cdot \frac{1}{\ln(1-p_{\underline{\zeta}}(d_2))^{-1}} \left| \frac{\ln d_1 - \ln \zeta_1}{\ln d_2 - \ln d_1} \right| \right]^2
\end{aligned}$$

Hence if

$$\underline{\zeta}^0 = \begin{pmatrix} \zeta_1^0 \\ 0 \\ \zeta_2^0 \end{pmatrix} = \begin{pmatrix} d^0(p_0) \\ \\ d^0(p_t) \end{pmatrix} = \underline{c} ,$$

a given vector

$$\underline{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} ,$$

and the parameter $\underline{\zeta}$ is known to be close to $\underline{c}$, then the asymptotic standard deviation of $\hat{\underline{\zeta}}_1$ based on a sample of size N is:

$$\sqrt{\frac{x^{11}(\underline{\zeta})}{N}} \Big|_{\underline{\zeta}=\underline{c}}$$

$$= \frac{c_1(\ln c_2 - \ln c_1)}{\sqrt{N} (\ln \ln(1-p_t))^{-1} - \ln \ln(1-p_0)^{-1}}$$

$$\times \left[\sqrt{\frac{p^0(d_1)}{1-p^0(d_1)}} \frac{1}{\ln(1-p^0(d_1))^{-1}} \left| \frac{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-p_0)^{-1}}{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-p^0(d_1))^{-1}} \right| \right. \\ \left. + \sqrt{\frac{p^0(d_2)}{1-p^0(d_2)}} \frac{1}{\ln(1-p^0(d_2))^{-1}} \left| \frac{\ln \ln(1-p^0(d_1))^{-1} - \ln \ln(1-p_0)^{-1}}{\ln \ln(1-p^0(d_2))^{-1} - \ln \ln(1-p^0(d_1))^{-1}} \right| \right]$$

The part of the expression to be minimized is the same as that for estimating the low dose response rate if we are making inference at the same point $(d_0, p^0(d_0))$ or $(d^0(p_0), p_0)$. Hence the following results hold:

- (1) the optimal design in the observable region is at

$$d_1 = d_1(\underline{\zeta})|_{\underline{\zeta}=\underline{c}}, d_2 = d_2(\underline{\zeta})|_{\underline{\zeta}=\underline{c}} \text{ where } (p^0(d_1(\underline{\zeta})), p^0(d_2(\underline{\zeta}))|_{\underline{\zeta}=\underline{c}} = (.1, .99);$$

- (2) the surfaces of $\sqrt{x^{11}(\underline{\zeta})/N}|_{\underline{\zeta}=\underline{c}}$ are fairly flat near the optimum so that experimentation need not be performed at exactly the optimal design,

- (3) the optimal allocating fraction $\rho_1(\underline{\zeta})|_{\underline{\zeta}=\underline{c}}$ is about 60-70%.

Hence for both cases of the model with β known or unknown, the locally optimal design for estimating the low dose response rate and the safe dose are the same.

CHAPTER III

A SMALL SAMPLE INVESTIGATION OF THE USE OF DIFFERENT BERNOULLI ESTIMATORS FOR LOW DOSE EXTRAPOLATION

3.1. Motivation

The following example motivated the investigation in this chapter.

The experiment observes $U_1, \dots, U_n$ a random sample from a Bernoulli distribution, i.e., $U_1, \dots, U_n$ independent $\sim b(1, p)$, $0 < p < 1$, and the object is to use the data to estimate p .

Different estimators may be used:

if we use the method of maximum likelihood, the estimator is

$$\bar{U} = \frac{\sum_{i=1}^n U_i}{n} ;$$

if we use a beta (α, β) prior and square error loss, the Bayes estimator is

$$\frac{\alpha + \sum_{i=1}^n U_i}{\alpha + \beta + n} ,$$

and in particular, the Bayes estimator with a beta $(\sqrt{n}/2, \sqrt{n}/2)$ prior and square error loss is the minimax estimator

$$\frac{\bar{U} + 1/(2\sqrt{n})}{1 + 1/\sqrt{n}}$$

The risks (with square error loss) for these estimators are:

$$R(p, \bar{U}) = \frac{p(1-p)}{n} \quad \text{and} \quad R\left(p, \frac{\bar{U} + 1/(2\sqrt{n})}{1 + 1/\sqrt{n}}\right) = \frac{1}{4(1 + \sqrt{n})^2}.$$

For $n \geq 1$,

Diagram 1

The graphs of Diagram 1 show that the minimax estimator, with the contraction of $\bar{U}$ towards $1/2$, not only estimates better (in terms of risk with square error loss or mean square error) when experimentation is performed at $p = 1/2$ but also for a neighborhood of $1/2$. The risk functions cross over at two points

$$p = \frac{1}{2} \pm \sqrt{\frac{1}{4} \left(1 - \frac{n}{(1 + \sqrt{n})^2} \right)}$$

and both converge to $1/2$ as $n \rightarrow \infty$. Hence if risk with square error loss is the criterion for choice of estimators for p , the estimator preferred may depend on the region of experimentation. Even though the maximum likelihood estimator is asymptotically as good as the minimax estimator, for finite and especially small sample size, if experimentation is performed in the middle region (a neighborhood of $1/2$), the minimax rather than the maximum likelihood estimator provides smaller mean square error.

The example bears certain relevance to our low dose extrapolation problem:

1. we are taking Bernoulli data at different dose levels;
2. we are forced to experiment in the middle region where the probability of response at each dose level is observable;
3. the Bernoulli parameters at the experimental dose levels are natural parameters without any model assumption, and inferences on the low dose effects may be regarded as inferences on these parameters for estimating the function in the assumed model for extrapolation;

4. maximum likelihood estimators are often used for estimations related to dose response functions;
5. the small sample situation may arise if the experimental subjects are more precious than mice and few subjects are available for experimentation.

Thus the following sections are aimed at investigating, with small sample sizes and for the purpose of low dose extrapolation, how different estimators of the experimental Bernoulli parameters behave in different regions of experimentation. The purpose is to see if there is a different preference in estimator for low dose extrapolation when we are experimenting in the middle region with small sample sizes. In particular, because of the small sample size used, the 'observable' region will be taken to have response rate between .1 and .9 (rather than .001 and .999).

3.2. The One Parameter Case

Assume $p^0(d) \in P_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty\}$, $0 < \beta < \infty$ known. Given $0 < d_0 < b_1$, the object is to estimate $p^0(d_0)$.

For comparison with the last chapter, we will consider a one point design with N subjects exposed to a dose d_1 . Since X_1 (the sum of the dichotomous responses) $\sim b(N, p(d_1))$, $0 < p(d_1) < 1$, is sufficient for $p(d_1)$, we will only consider estimators of $p(d_1)$ based on X_1 .

By an argument similar to that in Section 2.3, Case 1, $\bar{P}_1 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 \leq \theta \leq \infty\}$, $0 < \beta < \infty$ known, can be

reparametrized as $\bar{P}_1 = \{p_{p(d_1)}(d) = 1 - (1 - p(d_1))^{(d/d_1)^\beta}, 0 \leq d < \infty: 0 \leq p(d_1) \leq 1\}$.

The one-one correspondence between the two parametrizations is

$$\theta = \frac{-\ln(1 - p(d_1))}{d_1^\beta}.$$

Hence if $\widetilde{p}(d_1)$ is some estimator of $p(d_1)$ based on X_1 , and $\widetilde{p}(d_1)_x$ is the estimate when $X_1 = x$ is observed, then $p(d_0)$ can be estimated in the following way:

1. estimate $p(d_1)$ by $\widetilde{p}(d_1)_x$,
2. estimate the function in $\bar{P}_1$ by requiring it to pass through $(d, p(d)) = (d_1, \widetilde{p}(d_1)_x)$, i.e. estimate the parameter of the model $p(d_1)$ by $\widetilde{p}(d_1)_x$ or θ by

$$\tilde{\theta}_x = \frac{-\ln(1 - \widetilde{p}(d_1)_x)}{d_1^\beta},$$

3. finally, estimate $p(d_0)$ by extrapolating the estimated function to d_0 , i.e., estimate $p(d_0)$ by

$$1 - (1 - \widetilde{p}(d_1)_x)^{(d_0/d_1)^\beta} \quad \text{or} \quad 1 - \exp(-\tilde{\theta}_x d_0^\beta).$$

The procedure defines an estimator for $p(d_0)$; let it be denoted by $\widetilde{p}(d_0)$. Since it is based on X_1 , let $\widetilde{p}(d_0)_x$ be the estimate when $X_1 = x$ is observed. In particular, if $\widetilde{p}(d_1)_x = p^0(d_1)$, then $\widetilde{p}(d_0)_x = p^0(d_0)$. Thus the effectiveness of the estimator $\widetilde{p}(d_0)$ for $p(d_0)$ depends closely on the estimator $\widetilde{p}(d_1)$ for $p(d_1)$.

We will evaluate the effectiveness of an estimator $\widetilde{p}(d_0)$ of $p(d_0)$ by its mean square error. The bias of

$$\begin{aligned}\widetilde{p}(d_0) &= b(\widetilde{p}(d_0)) \\ &= E_{p^0(d_1)}[\widetilde{p}(d_0)] - p^0(d_0) . \\ &= \sum_{x=0}^N \binom{N}{x} p^0(d_1)^x (1-p^0(d_1))^{N-x} \widetilde{p}(d_0)_x - p^0(d_0),\end{aligned}$$

the standard deviation of $\widetilde{p}(d_0)$

$$\begin{aligned}&= \sigma(\widetilde{p}(d_0)) \\ &= \sqrt{E_{p^0(d_1)}[\widetilde{p}(d_0) - E_{p^0(d_1)}\widetilde{p}(d_0)]^2} \\ &= \sqrt{\sum_{x=0}^N \binom{N}{x} p^0(d_1)^x (1-p^0(d_1))^{N-x} (\widetilde{p}(d_0)_x - E_{p^0(d_1)}\widetilde{p}(d_0))^2}\end{aligned}$$

and the square root of the mean square error of $\widetilde{p}(d_0)$

$$\begin{aligned}&= \sqrt{\text{MSE}(\widetilde{p}(d_0))} \\ &= \sqrt{[b^2(\widetilde{p}(d_0)) + \sigma^2(\widetilde{p}(d_0))]} \end{aligned}$$

The three quantities $b(\widetilde{p}(d_0))$, $\sigma(\widetilde{p}(d_0))$, and $\sqrt{\text{MSE}(\widetilde{p}(d_0))}$ depend upon the sample size N , the design $p^0(d_1)$, the estimator $\widetilde{p}(d_1)$, the model used for extrapolation, and the inference dose d_0 .

We will examine two estimators here:

1. the maximum likelihood estimator

$$\hat{p}(d_0) = 1 - (1 - \hat{p}(d_1))^{(d_0/d_1)^\beta}$$

with the Bernoulli maximum likelihood estimator $\hat{p}(d_1) = X_1/N$, and

2.
$$\check{p}(d_0) = 1 - (1 - \check{p}(d_1))^{(d_0/d_1)^\beta}$$

with the Bernoulli minimax estimator

$$\check{p}(d_1) = \frac{X_1/N + 1/(2\sqrt{N})}{1 + 1/\sqrt{N}}$$

Several different cases for inference at low dose are included in the numerical investigation:

$p^0(d)$ through the point $(d=z_1, p^0(d)=z_2)$ where (z_1^β, z_2) is as indicated		sample size N	(inference dose) $^\beta$ $= d_0^\beta$	$p^0(d_0)$
(A)	$(10^{-6}, 10^{-5})$	10	10^{-6}	10^{-5}
(B)	$(10^{-6}, 10^{-8})$	10	10^{-6}	10^{-8}
(C)	$(10^{-6}, 10^{-8})$	30	10^{-6}	10^{-8}
(D)	$(10^{-6}, 10^{-8})$	10	10^{-7}	10^{-9}

Cases (A) and (B) are chosen to have different underlying dose response functions (defined by different inference level $p^0(d_0)$). Cases (C) and (D) have the same underlying dose response function as (B), with Case (C) examining the situation with more observations, and Case (D), that of a more distant extrapolation.

Figures 5-16 show the numerical results for these estimators and illustrative cases. Each figure covers both estimators $\widetilde{p}(d_0) = \widehat{p}(d_0)$ and $\widetilde{p}(d_0) = \check{p}(d_0)$. Figures 5-8 are plots of the bias $b(\widetilde{p}(d_0))$ versus designs $p^0(d_1)$ for Cases (A)-(D), Figures 9-12 are plots of $\sigma(\widetilde{p}(d_0))$ versus $p^0(d_1)$, and Figures 13-16 are plots of $\sqrt{\text{MSE}(\widetilde{p}(d_0))}$ versus $p^0(d_1)$. The scales for $|b(\widetilde{p}(d_0))|/p^0(d_0)$, $\sigma(\widetilde{p}(d_0))/p^0(d_0)$, and $\sqrt{\text{MSE}(\widetilde{p}(d_0))}/p^0(d_0)$ are shown on the right-hand margin of the appropriate graphs, permitting easy assessment of error relative to the target response rate being estimated. Three of the asymptotic curves in Figure 1 are reproduced in Figures 9, 10, and 11 for comparison.

FIGURE 5. Bias as it depends on the design point $p^0(d_1)$

$b(p(d_0))$ vs $p^0(d_1)$

CASE (A) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$$(z_1^\beta, z_2) = (10^{-6}, 10^{-5})$$

$$N = 10, d_0^\beta = 10^{-6}, p^0(d_0) = 10^{-5}$$

FIGURE 6. Bias as it depends on the design point $p^0(d_1)$

$b(\tilde{p}(d_0))$ vs $p^0(d_1)$

CASE (B) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$(z_1^B, z_2) = (10^{-6}, 10^{-8})$

$N = 10, d_0^B = 10^{-6}, p^0(d_0) = 10^{-8}$

FIGURE 7. Bias as it depends on the design point $p^0(d_1)$

$b(\widetilde{p}(d_0))$ vs $p^0(d_1)$

CASE (C) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$(z_1^\beta, z_2) = (10^{-6}, 10^{-8})$

$N = 30, d_0^\beta = 10^{-6}, p^0(d_0) = 10^{-8}$

FIGURE 8. Bias as it depends on the design point $p^0(d_1)$

$b(\tilde{p}(d_0))$ vs $p^0(d_1)$

CASE (D) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$$(z_1^B, z_2) = (10^{-6}, 10^{-8})$$

$$N = 10, d_0^B = 10^{-7}, p^0(d_0) = 10^{-9}$$

The following observations have been made:

Estimator from MLE of $p(d_1)$

$$\hat{p}(d_0)$$

Estimator from minimax

estimator of $p(d_1)$

$$\check{p}(d_0)$$

1. From graphs of $b(\hat{p}(d_0))$

For Cases (A), (B), (C), (D)

The bias $b(\hat{p}(d_0))$ increases monotonically from negative to positive values as the design $p^0(d_1)$ increases.

$\hat{p}(d_0)$ is unbiased when $p^0(d_1) = p^0(d_0)$, hence $b(\hat{p}(d_0))$ is positive for all $d_1 > d_0$ and is negative for $d_1 < d_0$.

$b(\hat{p}(d_0))$ increases to very large positive values for large $p^0(d_1)$ and exceeds $p^0(d_0)$ in order of magnitude for $p^0(d_1) \geq .32, .16, .55, .13$ in Cases (A)-(D). The negative bias cannot exceed $p^0(d_0)$ in order of magnitude but is close to such for very small $p^0(d_1)$.

$b(\check{p}(d_0))$ decreases monotonically from positive to negative values as $p^0(d_1)$ increases.

$\check{p}(d_0)$ is unbiased at .57, .57, .54, .57 for Cases (A)-(D), so $b(\check{p}(d_0))$ is positive for roughly the lower half and negative for the upper half of the design space.

$b(\check{p}(d_0))$ assumes very large positive values for small $p^0(d_1)$ and exceeds $p^0(d_0)$ in order of magnitude for $p^0(d_1) \leq .1, .1, .07, .1$ for the four cases. The negative bias is close to one order of magnitude for very large $p^0(d_1)$.

For cases with the same $p^0(d)$ but different N :

Case (B): $N = 10$, Case (C): $N = 30$

With the sample size increasing from 10 to 30 there is a great improvement for medium $p^0(d_1)$.

With N increasing from 10 to 30, the difference is slight over the entire range of $p^0(d_1)$.

Comparison of $b(\hat{p}(d_0))$ and $b(\check{p}(d_0))$

- (1) The two estimators have their bias graphs positive and negative on opposite sides of the design space.
- (2) For inference at small d_0 ,
 - (a) in terms of absolute bias, $\hat{p}(d_0)$ is generally better for small $p^0(d_1)$, $\check{p}(d_0)$ is generally better for medium to large $p^0(d_1)$,
 - (b) with increasing sample size, $b(\hat{p}(d_0))$ appears to diminish at a fast rate while $b(\check{p}(d_0))$ differs only slightly.

FIGURE 9. Standard deviation as it depends on the design point $p^0(d_1)$

$\sigma(\tilde{p}(d_0))$ vs $p^0(d_1)$

CASE (A) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$(z_1^B, z_2) = (10^{-6}, 10^{-5})$

$N = 10, d_0^B = 10^{-6}, p^0(d_0) = 10^{-5}$

FIGURE 10. Standard deviation as it depends on the design point $p^0(d_1)$

$\sigma(\tilde{p}(d_0))$ vs $p^0(d_1)$

CASE (B) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$$(z_1^B, z_2^B) = (10^{-6}, 10^{-8})$$

$$N = 10, d_0^B = 10^{-6}, p^0(d_0) = 10^{-8}$$

FIGURE 11. Standard deviation as it depends on the design point $p^0(d_1)$

$$\sigma(p(\tilde{d}_0)) \text{ vs } p^0(d_1)$$

CASE (C) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$$(z_1^B, z_2) = (10^{-6}, 10^{-8})$$

$$N = 30, d_0^B = 10^{-6}, p^0(d_0) = 10^{-8}$$

FIGURE 12. Standard deviation as it depends on the design point $p^0(d_1)$

$\sigma(\widetilde{p(d_0)})$ vs $p^0(d_1)$

CASE (D) $p^0(d)$ through $(d, p^0(d)) = (z_1, z_2)$ where

$(z_1^\beta, z_2^\beta) = (10^{-6}, 10^{-8})$

$N = 10, d_0^\beta = 10^{-7}, p^0(d_0) = 10^{-9}$

2. From graphs of $\sigma(p(\check{d}_0))$

For Cases (A), (B), (C), (D)

The standard deviation $\sigma(p(\hat{d}_0))$ has a characteristic $\sim$ shape.

$\sigma(p(\hat{d}_0))$ is large for almost the entire range of $p^0(d_1)$ explored: for $N = 10$, the standard deviation exceeds $p^0(d_0)$ by more than one order of magnitude for the entire range of the design space investigated, for $N = 30$, it also exceeds by one order of magnitude except for designs with $p^0(d_1)$ in the short range from .04 to .29.

$\sigma(p(\check{d}_0))$ has a different characteristic shape $\sim$.

$\sigma(p(\check{d}_0))$ is small relative to the target $p^0(d_0)$ for medium to large $p^0(d_1)$. It exceeds by more than one order of magnitude for designs $p^0(d_1)$ in the range $10^{-6} - .08$ for Case (A), $10^{-9} - .08$ for Case (B), $10^{-10} - .02$ for Case (C), and $10^{-10} - .08$ for Case (D).

For cases with the same $p^0(d)$ but different N :

Case (B): $N = 10$, Case (C): $N = 30$

There is great improvement (as N increases from 10 to 30) for medium $p^0(d_1)$. The asymptotic curves of Figure 1 are redrawn here for comparison. For sample sizes

There is only slight improvement (as N increases from 10 to 30) over the entire range of the design space. For N as small as 10, 30, $\sigma(p(\check{d}_0))$ is already rather close to the

as small as 10, 30, $\sigma(\hat{p}(d_0))$ is very different from the asymptotic standard deviation curve with the same N , but $\sigma(\hat{p}(d_0))$ approaches the asymptotic expression at a fast rate as N increases from 10 to 30.

asymptotic standard deviation curves with the same N .

Comparison of $\sigma(\hat{p}(d_0))$ and $\sigma(\check{p}(d_0))$

- (1) The standard deviation $\sigma(\hat{p}(d_0))$ is large in general and $\sigma(\check{p}(d_0))$ is small in general. $\check{p}(d_0)$ is better in standard deviation for the entire range of the design space than $\hat{p}(d_0)$; the improvement is greatest for medium to large $p^0(d_1)$.
- (2) $\sigma(\hat{p}(d_0))$ is far from the asymptotic standard deviation curve for small sample sizes but the convergence to the asymptotic expression is fast. $\sigma(\check{p}(d_0))$ is already close to the asymptotic expression for small sample sizes.

FIGURE 13. $\sqrt{\text{MSE}}$ as it depends on the design point $p^0(d_1)$

$\sqrt{\text{MSE}(\tilde{p}(d_0))}$ vs $p^0(d_1)$

CASE (A) $p^0(d)$ through (z_1, z_2) where $(z_1^\beta, z_2^\beta) = (10^{-6}, 10^{-5})$

$N = 10, d_0^\beta = 10^{-6}, p^0(d_0) = 10^{-5}$

FIGURE 14. $\sqrt{\text{MSE}}$ as it depends on the design point $p^0(d_1)$

$\sqrt{\text{MSE}(\tilde{p}(d_0))}$ vs $p^0(d_1)$

CASE (B) $p^0(d)$ through (z_1, z_2) where $(z_1^\beta, z_2^\beta) = (10^{-6}, 10^{-8})$

$N = 10, d_0^\beta = 10^{-6}, p^0(d_0) = 10^{-8}$

FIGURE 15. $\sqrt{\text{MSE}}$ as it depends on the design point $p^0(d_1)$

$\sqrt{\text{MSE}(\tilde{p}(d_0))}$ vs $p^0(d_1)$

CASE (C) $p^0(d)$ through (z_1, z_2) where $(z_1^\beta, z_2) = (10^{-6}, 10^{-8})$

$N = 30, d_0^\beta = 10^{-6}, p^0(d_0) = 10^{-8}$.

FIGURE 16. $\sqrt{\text{MSE}}$ as it depends on the design point $p^0(d_1)$

$\sqrt{\text{MSE}(\tilde{p}(d_0))}$ vs $p^0(d_1)$

CASE (D) $p^0(d)$ through (z_1, z_2) where $(z_1^\beta, z_2^\beta) = (10^{-6}, 10^{-8})$

$N = 10, d_0^\beta = 10^{-7}, p^0(d_0) = 10^{-9}$

3. From graphs of $\sqrt{\text{MSE}(\hat{p}(d_0))}$

For Cases (A), (B), (C), (D)

$\sqrt{\text{MSE}(\hat{p}(d_0))}$ has the characteristic shape $\sim$ resembling that of the standard deviation curve for $\sigma(\hat{p}(d_0))$ except for the rising (rather than falling) tail at large values of $p^0(d_1)$.

If we use $\sqrt{\text{MSE}(\hat{p}(d_0))}$ as the criterion for a design, the optimal design points are .08, .02, .24, .01 for Cases (A)-(D). All are very far from the asymptotic optimal design of .8.

A sample size of 10 is not adequate to estimate $p^0(d_0)$ to the right order of magnitude: $\sqrt{\text{MSE}(\hat{p}(d_0))}$ exceeds $p^0(d_0)$ by more than one order of magnitude for the entire design space. A sample size of 30 is only adequate for a short range $.04 \leq p^0(d_1) \leq .29$ near the optimal.

$\sqrt{\text{MSE}(\hat{p}(d_0))}$ increases to very large values for large $p^0(d_1)$.

$\sqrt{\text{MSE}(\check{p}(d_0))}$ has the characteristic shape $\cap$ resembling that of $\sigma(\check{p}(d_0))$ for medium $p^0(d_1)$ and that of $b(\check{p}(d_0))$ for small and large $p^0(d_1)$.

If we use $\sqrt{\text{MSE}(\check{p}(d_0))}$ as the criterion for a design, the optimal design points are .81, .81, .75, .81 for Cases (A)-(D). All are rather close to the asymptotic value of .8.

A sample size of 10 or more is adequate to estimate $p^0(d_0)$ to the right order of magnitude for a wide range of designs: $\sqrt{\text{MSE}(\check{p}(d_0))}$ exceeds $p^0(d_0)$ by more than one order of magnitude only for $p^0(d_1) \leq .14, .14, .08, .14$ for Cases (A)-(D).

$\sqrt{\text{MSE}(\check{p}(d_0))}$ is large for small $p^0(d_1)$ but small for the rest of the design space.

For cases with the same $p^0(d)$ but different N :

Case (B): $N = 10$, Case (C): $N = 30$.

There is great improvement
(as N increases from 10 to 30)
for medium $p^0(d_1)$.

There is only slight improvement
(as N increases from 10 to 30) over
the entire range of the design space.

Comparison of $\sqrt{\text{MSE}(p(\hat{d}_0))}$ and $\sqrt{\text{MSE}(p(\check{d}_0))}$

- (1) $\sqrt{\text{MSE}(p(\hat{d}_0))}$ is the smaller for small $p^0(d_1)$, $\sqrt{\text{MSE}(p(\check{d}_0))}$ is the smaller for medium to large $p^0(d_1)$. The cross over points are .1, .03, .25, .02 for Cases (A)-(D).
- (2) The optimal designs (with respect to $\sqrt{\text{MSE}(p(\check{d}_0))}$ of $p(\hat{d}_0)$ are very different from the asymptotic optimal design while those of $p(\check{d}_0)$ are close to the asymptotic optimal design.

There is in here certain similarity to the example in Section 3.1. Recall that for a random sample from a $b(1, p)$ distribution, the minimax estimator of p is better in mean square error than the maximum likelihood estimator of p for p near .5. In our one parameter case (which includes all the hit models), when the inference level $p^0(d_0)$ is small, this superiority in meansquare error propagates to the extrapolated estimator of $p(d_0)$. We shall show that for inference at a low dose, the improvement in mean square error of $p(\check{d}_0)$ over $p(\hat{d}_0)$ in a neighborhood of $p^0(d_1) = .5$ hold for more general one parameter models.

The situation can be explained by the following diagram for $N = 4$, but it holds for $N \geq 1$.

Diagram 2

The solid curve in Diagram 2 is $p^0(d)$. When the maximum likelihood estimator $\hat{p}(d_0)$ is used, the estimation procedure is as follows: the points marked x represent the possible values of X_1/N and the ---- lines are the functions for extrapolation. The values of $\hat{p}(d_0)$ are marked $\bullet$. The estimation procedure for $\check{p}(d_0)$ is similar. When experimenting at $d_1 = d^0(1/2)$, i.e. $p^0(d_1) = .5$, the contraction of x towards .5 by $(X_1/N + 1/(2\sqrt{N})) / (1 + 1/\sqrt{N})$ implies that the curves are to be used for extrapolation instead, and the values of $\check{p}(d_0)$ are marked by the circles o . It is obvious that $\check{p}(d_0)$ will have better mean

square error than $\hat{p}(d_0)$ when experimenting at $d_1 = d^0(1/2)$. Hence by continuity, there is a neighborhood $\mathcal{O}$ of $d^0(1/2)$ such that $\check{p}(d_0)$ is better than $\hat{p}(d_0)$ if experimentation is performed at a point in $\mathcal{O}$.

The only requirement on the one parameter model is that on reparametrizing $p(d)$ in terms of $p(d^0(1/2))$ (the parameter for the response rate at dose $d^0(1/2)$), the functional value at d_0 increases with $p(d^0(1/2))$. This condition holds for the probit and logit models with β known, holds for P_1 or a model with d^β in P_1 replaced by any other metameter $g(d):[0,\infty) \rightarrow [0,\infty)$ known monotone increasing one-one onto function when inference is to be made on a low dose. The contraction towards .5 may be interpreted as providing a more precise estimator for that response rate but it is shown here that the contraction is effective for a wide neighborhood of .5.

The results indicate that if we are making inference at a low dose and we use $\sqrt{\text{MSE}(\check{p}(d_0))}$ as the criterion to choose our estimator, different estimators may behave differently in different regions of experimentation. For small sample sizes, the two estimators $\hat{p}(d_0)$ and $\check{p}(d_0)$ we investigated have different behaviors: $\hat{p}(d_0)$ derived from the Bernoulli maximum likelihood estimator $\hat{p}(d_1)$, also a maximum likelihood estimator itself, is preferred when experimenting at small $p^0(d_1)$, $\check{p}(d_0)$ derived from the Bernoulli minimax estimator $\check{p}(d_1)$ is preferred when experimenting at medium to large $p^0(d_1)$. For large sample sizes, $\check{p}(d_0) \xrightarrow{\text{a.s.}} \hat{p}(d_0)$ and the two estimators are asymptotically the same. Since $\check{p}(d_0)$ is better in $\sqrt{\text{MSE}(\check{p}(d_0))}$ than $\hat{p}(d_0)$ for all sample sizes

when experimenting at d_1 where $p^0(d_1) = .5$, the cross over point cannot exceed .5. Hence $\check{p}(\check{d}_0)$, with its better mean square error in the upper half of the design space and a more stable optimal design, appears to be more suitable than $\hat{p}(\hat{d}_0)$ for low dose extrapolation when experimentation, with a single point design and a small sample size, is performed in the preferred observable region.

3.3. The Two Parameter Case

Assume $p^0(d) \in P_2 = \{p_\theta(d) = 1 - e^{-\theta d^\beta}, 0 \leq d < \infty; 0 < \theta < \infty, 0 < \beta < \infty\}$. Given $0 < d_0 < b_1$, the object is to estimate $p^0(d_0)$.

We will consider a two point design with $n_i = \rho_i N$ subjects exposed to dose d_i , $i = 1, 2$, $0 < d_1 < d_2$ and $X_i \sim b(n_i, p(d_i))$, $i = 1, 2$ are observed, $0 < p(d_1) < p(d_2) < 1$.

From (2.3.6), the model

$$P_2 = \left\{ p_{\left(\begin{smallmatrix} p(d_1) \\ p(d_2) \end{smallmatrix} \right)}(d) = 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - p(d_1)} + \frac{\ln d - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - p(d_2)} \right] \right), \right. \\ \left. 0 \leq d < \infty; 0 < p(d_1) < p(d_2) < 1 \right\}.$$

Let $\widetilde{p}(d_i)$ be some estimator of $p(d_i)$ based on X_i , $i = 1, 2$ and $\widetilde{p}(d_i)_{x_i}$ be the estimate when $X_i = x_i$ is observed, $i = 1, 2$. If $0 < \widetilde{p}(d_1)_{x_1} < \widetilde{p}(d_2)_{x_2} < 1$, then $p(d_0)$ is to be estimated by a procedure similar to that for the one parameter case:

1. estimate $p(d_i)$ by $p(\widetilde{d}_i)_{x_i}$, $i = 1, 2$,
2. estimate the function in P_2 by requiring it to pass through $(d_i, p(\widetilde{d}_i)_{x_i})$, $i = 1, 2$,
3. estimate $p(d_0)$ by extrapolating the estimated function to d_0 .

The estimator is extended to the whole sample space in parallel with that in Section 2.3 Case 2. The resultant estimator is as follows:

$$\widetilde{p}(\widetilde{d}_0) = \begin{cases} 1 - \exp \left(-\exp \left[\frac{\ln d_2 - \ln d_0}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - p(\widetilde{d}_1)} + \frac{\ln d_0 - \ln d_1}{\ln d_2 - \ln d_1} \ln \ln \frac{1}{1 - p(\widetilde{d}_2)} \right] \right), & \text{if } 0 < p(\widetilde{d}_1) \leq p(\widetilde{d}_2) < 1 \\ \frac{n_1 p(\widetilde{d}_1) + n_2 p(\widetilde{d}_2)}{n_1 + n_2}, & \text{if } 0 \leq p(\widetilde{d}_2) < p(\widetilde{d}_1) \leq 1 \\ & \text{or } p(\widetilde{d}_1) = p(\widetilde{d}_2) = 1 \\ 0, & \text{if } p(\widetilde{d}_1) = 0, 0 \leq p(\widetilde{d}_2) \leq 1 \\ & \text{or } 0 < p(\widetilde{d}_1) < 1, p(\widetilde{d}_2) = 1. \end{cases}$$

Let $\widetilde{p}(\widetilde{d}_0)_{\underline{x}}$ be the estimate when

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \underline{x}$$

is observed. If $p(\widetilde{d}_1)_{\underline{x}} = p^0(d_1)$, $p(\widetilde{d}_2)_{\underline{x}} = p^0(d_2)$ then $\widetilde{p}(\widetilde{d}_0)_{\underline{x}} = p^0(d_0)$; hence the effectiveness of the estimator $\widetilde{p}(\widetilde{d}_0)$ for $p(d_0)$ depends on the estimators $p(\widetilde{d}_1)$, $p(\widetilde{d}_2)$ for $p(d_1)$, $p(d_2)$.

The mean square error is again used as the criterion to evaluate an estimator $\widetilde{p}(d_0)$ of $p(d_0)$. The two components of mean square error are also examined. The bias of

$$\begin{aligned} \widetilde{p}(d_0) &= b(\widetilde{p}(d_0)) \\ &= E \left(\begin{pmatrix} p^0(d_1) \\ p^0(d_2) \end{pmatrix} \right) \widetilde{p}(d_0) - p^0(d_0) \\ &= \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} \binom{n_2}{x_2} p^0(d_1)^{x_1} (1-p^0(d_1))^{n_1-x_1} p^0(d_2)^{x_2} (1-p^0(d_2))^{n_2-x_2} \\ &\quad \cdot \widetilde{p}(d_0)_{\underline{x}} - p^0(d_0), \end{aligned}$$

the standard deviation of $\widetilde{p}(d_0)$

$$\begin{aligned} &= \sigma(\widetilde{p}(d_0)) \\ &= \sqrt{\left[\sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} \binom{n_2}{x_2} p^0(d_1)^{x_1} (1-p^0(d_1))^{n_1-x_1} p^0(d_2)^{x_2} \right. \\ &\quad \cdot (1-p^0(d_2))^{n_2-x_2} \left. \left(\widetilde{p}(d_0)_{\underline{x}} - E \left(\begin{pmatrix} p^0(d_1) \\ p^0(d_2) \end{pmatrix} \right) \widetilde{p}(d_0) \right)^2 \right]}, \end{aligned}$$

and the square root of the mean square error of $\widetilde{p}(d_0)$

$$\begin{aligned} &= \sqrt{\text{MSE}(\widetilde{p}(d_0))} \\ &= \sqrt{[b^2(\widetilde{p}(d_0)) + \sigma^2(\widetilde{p}(d_0))]} . \end{aligned}$$

The three quantities $b(\widetilde{p}(d_0))$, $\sigma(\widetilde{p}(d_0))$ and $\sqrt{\text{MSE}}(\widetilde{p}(d_0))$ depend upon the sample size N , the allocation fractions ρ_i , $i = 1, 2$, the design $(p^0(d_1), p^0(d_2))$, the estimators $\widetilde{p}(d_1)$, $\widetilde{p}(d_2)$, the model for extrapolation, the true dose response function $p^0(d)$, and the inference dose d_0 .

We will investigate how $\widetilde{p}(d_0)$ behave in different regions of experimentation for different Bernoulli estimators $\widetilde{p}(d_1)$ and $\widetilde{p}(d_2)$. One set of Bernoulli estimators to examine is $\widehat{p}(d_1) = \bar{X}_1 = X_1/n_1$, $\widehat{p}(d_2) = \bar{X}_2 = X_2/n_2$, but in order to avoid the effect of the arbitrary definition of $\widetilde{p}(d_0)$ when X_1/n_1 or X_2/n_2 of the observation equals zero or one, we will examine $\widehat{\widetilde{p}}(d_0)$ based on

$$\widehat{p}(d_1) = \frac{\bar{X}_1 + .1/n_1^8}{1 + 1/n_1^8}$$

and

$$\widehat{p}(d_2) = \frac{\bar{X}_2 + .99/n_2^8}{1 + 1/n_2^8}$$

as an approximation to the modified maximum likelihood estimator $\widehat{\widehat{p}}(d_0)$ based on $\widehat{p}(d_1)$ and $\widehat{p}(d_2)$. The numbers .1, .99 are used to make these estimators analogous to the next set of estimators and $\widehat{p}(d_1)$, $\widehat{p}(d_2)$ contract $\bar{X}_1$, $\bar{X}_2$ towards these points by a very small amount. The next set of Bernoulli estimators is

$$\widehat{\widehat{p}}(d_1) = \frac{\bar{X}_1 + .1/n_1^{1/8}}{1 + 1/n_1^{1/8}} \quad \text{and} \quad \widehat{\widehat{p}}(d_2) = \frac{\bar{X}_2 + .99/n_2^{1/8}}{1 + 1/n_2^{1/8}}.$$

There is strong pulling of $\bar{X}_1, \bar{X}_2$ towards the points .1 and .99. The choice of .1, .99 is motivated by the asymptotic optimality of the design $(p^0(d_1), p^0(d_2)) = (.1, .99)$ which suggest that it would be advantageous to pindown precisely at those response levels. We will refer to these estimators of $p(d_0)$ resulting from the two sets of Bernoulli estimators as $\widehat{p}(d_0)$ and $\check{p}(d_0)$.

A brief examination has been made for the following case:

$p^0(d)$ through the points	sample size		inference dose	
$(d, p^0(d)) = (z_{11}, z_{12}), (z_{21}, z_{22})$	n_1	n_2	d_0	$p^0(d_0)$
$(10^{-7}, 10^{-11}), (10^{-6}, 10^{-8})$	18	12	10^{-6}	10^{-8}

The allocation fraction has been chosen to approximate the asymptotic optimal fraction.

Tables 17-19 give the mean, standard deviation and square root of the mean square error for $\widehat{p}(d_0)$ for the different designs with response levels between 10^{-8} to $1-10^{-6}$. Tables 20-22 give the corresponding quantities for $\check{p}(d_0)$. The two axes of each table give the true response rates at the two experimental doses $p^0(d_1), p^0(d_2)$. Recall that we will take the response rates from .1 to .9 as observable, designs in the observable region thus lie to the lower right of the dotted lines. Each entry in the table is written in the same way as previously $(a_1 \cdot a_2 \times 10^{-b} \text{ as } -b \begin{smallmatrix} a_1 & a_2 \\ 1 & 2 \end{smallmatrix})$ and gives the quantity (mean, standard

TABLE 17 Expected value of $\hat{p}(d_0)$ for various designs ($p^0(d_1)$, $p^0(d_2)$) with $n_1 = 18$, $n_2 = 12$

Ideal value of $E \hat{p}(d_0)$ would be 10^{-8} (written as -8^{10}) throughout the table.

Lower Design Point $p^0(d_1)$

Upper Design Point $p^0(d_2)$

	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$
$1-10^{-8}$	10 -8	20 -7	11 -6	27 -7	51 -8	36 -8	32 -8	10 -7	39 -6	10 -4	16 -3	18 -2	15 -1	83 -1	10 0	78 -1						
$1-10^{-7}$																						
$1-10^{-6}$	10 -8	19 -7	92 -7	15 -7	17 -8	97 -9	99 -9	86 -8	39 -6	10 -4	16 -3	18 -2	15 -1	83 -1	79 -1							
$1-10^{-5}$																						
$1-10^{-4}$	10 -8	18 -7	79 -7	88 -8	25 -8	61 -7	18 -7	72 -6	66 -4	11 -3	17 -2	18 -1	15 -1	83 -1								
$1-10^{-3}$																						
$1-10^{-2}$	10 -8	18 -7	97 -7	47 -7	58 -7	15 -6	54 -6	26 -5	16 -4	98 -3	58 -2	33 -1	19 -1									
.9	10 -8	28 -7	30 -6	22 -6	41 -5	27 -4	20 -3	13 -3	63 -2	24 -2	77 -1	21 -1										
.8	10 -8	34 -7	43 -6	40 -6	22 -5	26 -4	19 -3	93 -3	33 -2	93 -1	22 -1											
.7	10 -8	37 -7	55 -6	74 -6	13 -4	15 -3	87 -3	32 -2	89 -1	20 -1												
.6	10 -8	41 -7	69 -6	16 -5	62 -4	58 -3	26 -2	77 -2	17 -1													
.5	10 -8	44 -7	90 -6	45 -5	23 -3	17 -2	59 -2	14 -1														
.4	10 -8	49 -7	12 -5	14 -4	70 -3	38 -2	11 -1															
.3	10 -8	55 -7	19 -5	42 -4	17 -2	70 -2																
.2	97 -9	63 -7	29 -5	11 -3	34 -2																	
.1	89 -9	65 -7	45 -5	26 -3																		
10^{-2}	65 -9	62 -7	58 -5																			
10^{-3}																						
10^{-4}	60 -9	60 -7																				
10^{-5}																						
10^{-6}	60 -9																					
10^{-7}																						

TABLE 18 Standard deviation of $p(d_0)$ for various designs ($p^0(d_1)$, $p^0(d_2)$) with $n_1 = 18$, $n_2 = 12$.

$p^0(d_0) = 10^{-8}$

Lower Design Point $p^0(d_1)$

	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$
$1-10^{-8}$	24 -5		46 -5		25 -5		12 -6	20 -7	54 -7	20 -5	26 -4	20 -3	10 -2	40 -2	13 -1	36 -1	37 -1		42 -2		33 -3	
$1-10^{-7}$																						
$1-10^{-6}$	24 -5		44 -5		22 -5		72 -7	51 -7	18 -6	21 -5	26 -4	20 -3	10 -2	40 -2	13 -1	36 -1	37 -1		34 -2			
$1-10^{-5}$																						
$1-10^{-4}$	24 -5		42 -5		19 -5		15 -6	31 -6	11 -5	47 -5	32 -4	21 -3	10 -2	41 -2	13 -1	36 -1	37 -1					
$1-10^{-3}$																						
$1-10^{-2}$	24 -5		45 -5		35 -5		79 -6	15 -5	74 -5	41 -4	19 -3	72 -3	23 -2	64 -2	17 -1	38 -1						
.9	24 -5		69 -5		83 -5		23 -5	20 -4	16 -3	73 -3	23 -2	59 -2	13 -1	23 -1	36 -1							
.8	24 -5		81 -5		11 -4		55 -5	12 -3	77 -3	27 -2	67 -2	14 -1	23 -1	34 -1								
.7	24 -5		89 -5		14 -4		16 -4	42 -3	21 -2	58 -2	12 -1	21 -1	30 -1									
.6	24 -5		97 -5		18 -4		45 -4	10 -2	41 -2	98 -2	18 -1	26 -1										
.5	24 -5		11 -4		25 -4		11 -3	21 -2	68 -2	14 -1	21 -1											
.4	24 -5		12 -4		37 -4		24 -3	35 -2	96 -2	16 -1												
.3	23 -5		13 -4		57 -4		44 -3	52 -2	11 -1													
.2	23 -5		15 -4		84 -4		73 -3	65 -2														
.1	21 -5		17 -4		12 -3		11 -2															
10^{-2}	16 -5		15 -4		14 -3																	
10^{-3}																						
10^{-4}	14 -5		14 -4																			
10^{-5}																						
10^{-6}	14 -5																					
10^{-7}																						

Upper Design Point $p^0(d_2)$

TABLE 19 $\sqrt{\text{MSE of } \hat{p}(d)}$ for various designs ($p^0(d_1)$, $p^0(d_2)$) with $n_1 = 18$, $n_2 = 12$.
 $p^0(d_0) = 10^{-8}$

		Lower Design Point $p^0(d_1)$																							
		10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$		
Upper Design Point $p^0(d_2)$	$1-10^{-8}$	24 -5	46 -5	25 -5		12 -6	21 -7	54 -7	20 -5	26 -4	20 -3	10 -2	40 -2	13 -1	39 -1	91 -1	10 0	78 -1							
	$1-10^{-7}$																								
	$1-10^{-6}$	24 -5	44 -5	22 -5		73 -7	51 -7	18 -6	21 -5	26 -4	20 -3	10 -2	40 -2	13 -1	39 -1	91 -1	80 -1								
	$1-10^{-5}$																								
	$1-10^{-4}$	24 -5	42 -5	19 -5		15 -6	31 -6	11 -5	47 -5	32 -4	21 -3	10 -2	41 -2	13 -1	39 -1	91 -1									
	$1-10^{-3}$																								
	$1-10^{-2}$	24 -5	45 -5	35 -5		79 -6	15 -5	74 -5	41 -4	19 -3	72 -3	23 -2	64 -2	17 -1	42 -1										
	.9	24 -5	69 -5	83 -5		23 -5	20 -4	16 -3	73 -3	23 -2	59 -2	13 -1	25 -1	42 -1											
	.8	24 -5	81 -5	11 -4		55 -5	12 -3	77 -3	27 -2	67 -2	14 -1	25 -1	40 -1												
	.7	24 -5	89 -5	14 -4		16 -4	42 -3	21 -2	59 -2	13 -1	23 -1	36 -1													
	.6	24 -5	97 -5	18 -4		45 -4	10 -2	42 -2	10 -1	19 -1	31 -1														
	.5	24 -5	11 -4	25 -4		11 -3	21 -2	70 -2	15 -1	25 -1															
	.4	24 -5	12 -4	37 -4		24 -3	36 -2	10 -1	20 -1																
	.3	23 -5	13 -4	57 -4		44 -3	55 -2	13 -1																	
	.2	23 -5	15 -4	84 -4		74 -3	74 -2																		
	.1	21 -5	17 -4	12 -3		11 -2																			
	10^{-2}	16 -5	15 -4	14 -3																					
	10^{-3}																								
	10^{-4}	14 -5	14 -4																						
	10^{-5}																								
	10^{-6}	14 -5																							
	10^{-7}																								

TABLE 20 Expected value of $p(d_0)$ for various designs ($p^0(d_1)$, $p^0(d_2)$) with $n_1 = 18$, $n_2 = 12$.
Ideal values of $E p(d_0)$ would be 10^{-8} (written as -8^{10}) throughout the table.

Lower Design Point $p^0(d_1)$ →

	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$
$1-10^{-8}$	41 -2		11 -2		11 -3		12 -5	16 -6	13 -6	13 -6	12 -6	12 -6	10 -6	79 -7	45 -7	11 -7	14 -10		16 -19		37 -33	
$1-10^{-7}$																						
$1-10^{-6}$	41 -2		11 -2		95 -4		75 -6	69 -7	44 -7	35 -7	29 -7	23 -7	16 -7	93 -8	33 -8	34 -9	60 -14		33 -29			
$1-10^{-5}$																						
$1-10^{-4}$	41 -2		10 -2		83 -4		45 -6	25 -7	12 -7	75 -8	49 -8	30 -8	16 -8	64 -9	15 -9	99 -11	47 -17					
$1-10^{-3}$																						
$1-10^{-2}$	41 -2		99 -3		70 -4		22 -6	72 -8	29 -8	18 -8	11 -8	71 -9	37 -9	13 -9	19 -10	30 -12						
.9	41 -2		11 -2		85 -4		28 -6	82 -8	31 -8	19 -8	14 -8	13 -8	19 -8	43 -8	11 -7							
.8	41 -2		12 -2		99 -4		32 -6	89 -8	37 -8	31 -8	51 -8	17 -7	77 -7	36 -6								
.7	41 -2		12 -2		11 -3		35 -6	93 -8	48 -8	78 -8	35 -7	22 -6	14 -5									
.6	41 -2		13 -2		12 -3		36 -6	91 -8	69 -8	28 -7	24 -6	19 -5										
.5	41 -2		13 -2		13 -3		35 -6	82 -8	11 -7	11 -6	14 -5											
.4	41 -2		14 -2		13 -3		31 -6	60 -8	21 -7	43 -6												
.3	41 -2		14 -2		14 -3		23 -6	29 -8	39 -7													
.2	41 -2		14 -2		14 -3		12 -6	63 -9														
.1	41 -2		14 -2		12 -3		16 -7															
10^{-2}	41 -2		11 -2		25 -4																	
10^{-3}																						
10^{-4}	41 -2		32 -3																			
10^{-5}																						
10^{-6}	41 -2																					
10^{-7}																						

↑ Upper Design Point $p^0(d_2)$

TABLE 21 Standard deviation of $p^0(d_0)$ for various designs ($p^0(d_1)$, $p^0(d_2)$) with $n_1 = 18$, $n_2 = 12$.
 $p^0(d_0) = 10^{-8}$

Upper Design Point $p^0(d_2)$

	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$
$1-10^{-8}$	14 -5	53 -5	86 -5	15 -5	37 -6	30 -6	28 -6	26 -6	24 -6	19 -6	13 -6	59 -7	10 -7	70 -11	13 -16	28 -17						
$1-10^{-7}$																						
$1-10^{-6}$	14 -5	53 -5	77 -5	10 -5	17 -6	11 -6	92 -7	74 -7	56 -7	37 -7	19 -7	71 -8	13 -8	87 -12	52 -17							
$1-10^{-5}$																						
$1-10^{-4}$	14 -5	85 -5	72 -5	67 -6	74 -7	42 -7	32 -7	28 -7	24 -7	20 -7	13 -7	50 -8	51 -9	23 -13								
$1-10^{-3}$																						
$1-10^{-2}$	14 -5	64 -4	16 -4	50 -6	59 -7	40 -7	35 -7	33 -7	31 -7	28 -7	22 -7	16 -7	14 -7									
.9	14 -5	13 -3	29 -4	74 -6	99 -7	10 -6	19 -6	48 -6	14 -5	39 -5	97 -5	21 -4										
.8	14 -5	12 -3	30 -4	91 -6	17 -6	32 -6	12 -5	50 -5	18 -4	52 -4	13 -3											
.7	14 -5	11 -3	31 -4	11 -5	28 -6	11 -5	59 -5	26 -4	89 -4	26 -3												
.6	14 -5	11 -3	33 -4	12 -5	48 -6	31 -5	20 -4	89 -4	30 -3													
.5	14 -5	11 -3	35 -4	13 -5	78 -6	78 -5	55 -4	25 -3														
.4	14 -5	11 -3	36 -4	13 -5	12 -5	17 -4	13 -3															
.3	14 -5	10 -3	36 -4	12 -5	16 -5	33 -4																
.2	14 -5	98 -4	34 -4	79 -6	20 -5																	
.1	14 -5	80 -4	26 -4	20 -6																		
10^{-2}	14 -5	29 -4	58 -5																			
10^{-3}																						
10^{-4}	14 -5	35 -5																				
10^{-5}																						
10^{-6}	14 -5																					
10^{-7}																						

Lower Design Point $p^0(d_1)$

TABLE 22 $\sqrt{\text{MSE of } p(d_1)}$ for various designs $(p^0(d_1), p^0(d_2))$ with $n_1 = 18, n_2 = 12$.
 $p^0(d_0) = 10^{-8}$

		Lower Design Point $p^0(d_1)$																									
		10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$	$1-10^{-7}$				
Upper Design Point $p^0(d_2)$	$1-10^{-8}$	41 -2		11 -2		11 -3		19 -5	41 -6	33 -6	31 -6	29 -6	26 -6	22 -6	15 -6	73 -7	14 -7	99 -9		10 -8		10 -8					
	$1-10^{-7}$																										
	$1-10^{-6}$	41 -2		11 -2		95 -4		13 -5	18 -6	12 -6	98 -7	79 -7	60 -7	40 -7	21 -7	75 -8	15 -8	10 -8		10 -8							
	$1-10^{-5}$																										
	$1-10^{-4}$	41 -2		10 -2		83 -4		81 -6	78 -7	43 -7	33 -7	28 -7	24 -7	20 -7	13 -7	50 -8	11 -8	10 -8									
	$1-10^{-3}$																										
	$1-10^{-2}$	41 -2		99 -3		72 -4		55 -6	59 -7	40 -7	35 -7	33 -7	31 -7	28 -7	22 -7	16 -7	14 -7										
	.9	41 -2		11 -2		90 -4		79 -6	10 -6	10 -6	19 -6	48 -6	14 -5	39 -5	97 -5	21 -4											
	.8	41 -2		12 -2		10 -3		97 -6	17 -6	32 -6	12 -5	50 -5	18 -4	52 -4	13 -3												
	.7	41 -2		12 -2		11 -3		11 -5	28 -6	11 -5	59 -5	26 -4	89 -4	26 -3													
	.6	41 -2		13 -2		12 -3		13 -5	48 -6	31 -5	20 -4	89 -4	30 -3														
	.5	41 -2		13 -2		13 -3		14 -5	78 -6	78 -5	55 -4	25 -3															
	.4	41 -2		14 -2		14 -3		14 -5	12 -5	17 -4	13 -3																
	.3	41 -2		14 -2		14 -3		12 -5	16 -5	33 -4																	
	.2	41 -2		14 -2		14 -3		80 -6	20 -5																		
	.1	41 -2		14 -2		12 -3		20 -6																			
	10^{-2}	41 -2		12 -2		25 -4																					
	10^{-3}																										
	10^{-4}	41 -2		32 -3																							
	10^{-5}																										
	10^{-6}	41 -2																									
	10^{-7}																										

deviation or $\sqrt{\text{MSE}}$) for that particular design consisting of experimenting at $p^0(d_1)$, $p^0(d_2)$ with the above allocation fraction ($\rho_1 = .6$) and sample size ($N = 30$).

The results show:

1. for $\widehat{p(d_0)}$ (from Tables 17-19):
 - (1) The bias is positive and importantly large for almost the entire region. The estimator is close to being unbiased when $p^0(d_1) = 10^{-8}$ or when the design $(p^0(d_1), p^0(d_2))$ is close to $(.3, 1-10^{-6})$.
 - (2) All three quantities (the bias, standard deviation and $\sqrt{\text{MSE}}$) assume large values (from 10^{-3} to 1) when both experimental levels are close to each other and have medium to high response rates (i.e. along the diagonal and the upper right vertex of the triangular grid).
 - (3) All these quantities (the bias, standard deviation and $\sqrt{\text{MSE}}$) are generally large for designs in the observable region (i.e. to the lower right of the dotted lines): the bias lies between 10^{-6} and 10^{-1} , the standard deviation and the $\sqrt{\text{MSE}}$ both lie between 10^{-4} and 10^{-1} .
 - (4) All three quantities (the bias, standard deviation and $\sqrt{\text{MSE}}$) are relatively small when $p^0(d_1)$ is small or when in combination with a large $p^0(d_2)$ (i.e. in the left part or the upper left corner of the table): the absolute bias lies between 0 and 10^{-5} , the standard deviation and the $\sqrt{\text{MSE}}$ lies between 10^{-7} and 10^{-4} .

- (5) The $\sqrt{\text{MSE}}$ exceeds the target value $p^0(d_0)$ in order of magnitude throughout the table. The standard deviation is the dominating component of the $\sqrt{\text{MSE}}$ for designs in the observable region. The optimal design (with respect to $\sqrt{\text{MSE}}$) within the observable region is at $(p^0(d_1), p^0(d_2)) = (.1, .9)$.

2. for $p(d_0)$ (from Tables 20-22).

- (1) The bias is large (from 10^{-4} to 10^{-2}) when $p^0(d_1)$ is small (in the left part of the table). It is relatively small (about 10^{-9} to 10^{-5}) for the rest of the design space.
- (2) The standard deviation is within one order of magnitude of the target value $p^0(d_0)$ when both experimental levels are large (the upper right vertex of the triangular grid) and ranges from 10^{-7} to 10^{-3} for the rest of the design space.
- (3) Of the two components of $\sqrt{\text{MSE}}$, the bias dominates when $p^0(d_1)$ is small or when both $p^0(d_1)$ and $p^0(d_2)$ are large, and the standard deviation dominates in the observable region.
- (4) The $\sqrt{\text{MSE}}$ exceeds the target value $p^0(d_0)$ in order of magnitude except for the design $(p^0(d_1), p^0(d_2)) = (1-10^{-2}, 1-10^{-8})$, and is close to one order of magnitude of $p^0(d_0)$ when both experimental levels are high. The optimal design in the observable region is $(.1, .9)$, followed closely by the design $(.2, .9)$.

If we compare the two estimators:

1. the approximate maximum likelihood estimator $\widehat{p}(d_0)$ has large positive bias when $p^0(d_1), p^0(d_2)$ are close to each other and both have medium to large values (i.e. along the diagonal and the upper right vertex of the triangular grid). The estimator $\widetilde{p}(d_0)$, using heavy contraction towards the points .1 and .99, reduces the bias by a great amount when the design point is near the upper right vertex of the grid and by a considerable amount when the design point is in the observable region. The bias for designs with small $p^0(d_1)$ (in the left part of the table) is highly increased.
2. the estimator $\widehat{p}(d_0)$ gives large standard deviation for designs along the diagonal or near the upper right vertex of the grid. The estimator $\widetilde{p}(d_0)$ reduces the standard deviation greatly for this part of the design space and considerably for the whole observable region.
3. $\widetilde{p}(d_0)$, compared with $\widehat{p}(d_0)$, increases the $\sqrt{\text{MSE}}$ for designs with small $p^0(d_1)$, reduces the $\sqrt{\text{MSE}}$ greatly for designs near the upper right vertex, and reduces the $\sqrt{\text{MSE}}$ by about a factor of 10^{-2} throughout the observable region.

Thus the two parameter case is similar to the example in Section 3.1 and the one parameter case in Section 3.2. In each of the cases, different estimators (for the corresponding quantity of interest) are preferred in different regions of experimentation. The Bayes estimator(s) of the Bernoulli parameter(s) contracts the maximum likelihood estimator(s) towards some centre (.5 in the example, .5 in the one parameter case and (.1, .99) in the two parameter case); the estimator derived through

such contraction(s) not only has smaller $\sqrt{\text{MSE}}$ (compared with the maximum likelihood estimator or approximate maximum likelihood estimator) when experimentation is performed at the response level(s) of the centre of contraction but also at response levels in the vicinity. For estimating the low dose response rate with small sample sizes: in the one parameter case, the maximum likelihood estimator is better at designs with low response rates and the contracted estimator is better at designs with medium to high response rates; in the two parameter case, the approximate maximum likelihood estimator $\widehat{p}(d_0)$ is better in the left part and upper left corner of the design space where $p^0(d_1)$ is small or in combination with a large $p^0(d_2)$, the contracted estimator $\widetilde{p}(d_0)$ is better in the observable region and when both experimental levels are high. Hence in either case, even though the maximum likelihood estimator and the contracted estimator are asymptotically equivalent, the contracted estimator may be preferable for low dose extrapolation if the experimentation is performed in the observable region with small sample sizes.

CHAPTER IV

DIFFERENTIATION OF MODELS WITH DIFFERENT LOW DOSE BEHAVIORS

4.1. Approach to the Model Testing Problem

The extreme value distribution model proposed in Section 2.1 for estimation of low dose behavior has the form

$$p(d) = 1 - e^{-\theta d^\beta}, \quad 0 < \theta < \infty, \quad 0 < \beta < \infty.$$

The model is chosen for its adequacy in the representation of behaviors in both the observable and the low dose ranges, and unlike the k-hit models, which fix the low dose behavior very discretely by the value of the integer k , the model allows the choice of the tail behavior to be made on a continuous scale. Since the model is a tolerance distribution model, and tolerance distribution models have been shown to have similar behaviors in the observable dose range but very different behaviors in the low dose range, the questions 'are we using the model with the right tail?' and 'which is the model with the right tail?' have to be considered. This chapter is an attempt to address such questions.

The problem with the choice of model is interpreted as follows: each of the tolerance distribution models is reasonably adequate in describing data in the experimental range on its own, but each model has its own tail behavior characteristic to the tail weight of its tolerance distribution (e.g. the probit model has its characteristic probit tail),

so that if we use any one of these tolerance distribution models for estimation of low dose behavior, we will have to know which model to use (e.g. the probit or the extreme value distribution model) to be able to estimate the tail behavior accurately.

Instead of trying to choose between different tolerance distribution models (e.g. to choose between the probit and the extreme value distribution models), which are all similar in the observable region, our approach is to take the extreme value distribution model and attach to it a full range of hit tail behaviors so as to preserve the low dose characteristics of the original model, to include the original model as a submodel, as well as to allow for the choice to be made for a different tail behavior.

The extended model proposed is as follows:

$$p(d) = 1 - \exp[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2})], \quad 0 \leq \theta_1 < \infty, 0 < \theta_2 < \infty, 0 < \beta_1 < \beta_2 < \infty,$$

and may be written (for simpler expressions in the next section) as:

$$p(d) = 1 - \exp \left[-(\xi_2 d)^{\xi_3} (1 + (\xi_1/d)^r)^{\xi_3} \right],$$

$$0 \leq \xi_1 < \infty, 0 < \xi_2 < \infty, 0 < \xi_3 < \infty, 0 < r < 1.$$

The correspondence in parameters are given by:

$$\theta_1 = \xi_1^r \xi_3 \xi_2^{\xi_3}, \quad \theta_2 = \xi_2^{\xi_3},$$

$$\beta_1 = (1-r)\xi_3, \quad \beta_2 = \xi_3.$$

Hence the extended model P_4

$$= \left\{ p(d) = 1 - \exp \left[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2}) \right], 0 \leq d < \infty: 0 \leq \theta_1 < \infty, 0 < \theta_2 < \infty, 0 < \beta_1 < \beta_2 < \infty \right\}$$

$$= \left\{ p(d) = 1 - \exp \left[-(\zeta_2 d)^{\zeta_3} (1 + (\zeta_1/d)^{\zeta_3}) \right], 0 \leq d < \infty: 0 \leq \zeta_1 < \infty, \right.$$

$$\left. 0 < \zeta_2 < \infty, 0 < \zeta_3 < \infty, 0 < \gamma < 1 \right\}.$$

With $\theta_1 = 0$ (or equivalently $\zeta_1 = 0$), the model P_4 reduces to the extreme value distribution model P_2 . The correspondence with the previous notation is

$$\theta = \theta_2 = \zeta_2^{\zeta_3}, \quad \beta = \beta_2 = \zeta_3.$$

For a dose response function in P_2 ,

$$p(d) \in P_2 \iff p(d) = 1 - \exp(-\theta_2 d^{\beta_2}), 0 \leq d < \infty,$$

where $\theta_2 > 0, \beta_2 > 0$ or

$$p(d) = 1 - \exp[-(\zeta_2 d)^{\zeta_3}], 0 \leq d < \infty,$$

where $\zeta_2 > 0, \zeta_3 > 0$

$$\iff \ln \ln \frac{1}{1-p(d)} = \ln \theta_2 + \beta_2 \ln d \quad (4.1.1)$$

or

$$\ln \ln \frac{1}{1-p(d)} = \zeta_3 \ln \zeta_2 + \zeta_3 \ln d \quad (4.1.2)$$

A plot of $\ln \ln[1-p(d)]^{-1}$ versus $\ln d$ will hence be a straight line with slope $\beta_2 (= \zeta_3)$ and intercept $\ln \theta_2 (= \zeta_3 \ln \zeta_2)$ as shown in Diagram 3.

The model P_2 consists of a family of straight lines of different slopes and intercepts in these $\ln \ln[1-p(d)]^{-1}$ versus $\ln d$ plots, and allows only two parameters (θ_2, β_2) (or (ζ_2, ζ_3)) for the description of behavior at all dose levels.

With $\theta_1 > 0$ (or equivalently $\zeta_1 > 0$), another set of parameters (θ_1, β_1) is allowed into the description. For a dose response function in $P_4 \setminus P_2$,

$$p(d) \in P_4 \setminus P_2 \iff p(d) = 1 - \exp[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2})], \quad 0 \leq d < \infty$$

where $\theta_1 > 0, \theta_2 > 0, \beta_1 > 0, \beta_2 > 0$ or

$$p(d) = 1 - \exp \left[-(\xi_2 d)^{\xi_3} (1 + (\xi_1/d)^{r\xi_3}) \right], \quad 0 \leq d < \infty,$$

where $\xi_1 > 0$, $\xi_2 > 0$, $\xi_3 > 0$, $0 < r < 1$.

$$\Leftrightarrow \ln \ln \frac{1}{1-p(d)} = \ln \theta_2 + \beta_2 \ln d + \ln \left[1 + \frac{\theta_1}{\theta_2} \left(\frac{1}{d} \right)^{\beta_2 - \beta_1} \right]$$

or

$$\ln \ln \frac{1}{1-p(d)} = \xi_3 \ln \xi_2 + \xi_3 \ln d + \ln \left[1 + (\xi_1/d)^{r\xi_3} \right] \quad (4.1.3)$$

Since $\beta_2 - \beta_1 (= r\xi_3) > 0$, for sufficiently large d ,

$$\ln \ln \frac{1}{1-p(d)} \doteq \ln \theta_2 + \beta_2 \ln d$$

or

$$\ln \ln \frac{1}{1-p(d)} \doteq \xi_3 \ln \xi_2 + \xi_3 \ln d \quad (4.1.4)$$

which are the same equations as (4.1.1) and (4.1.2); and for sufficiently small d ,

$$\frac{\theta_1}{\theta_2} \left(\frac{1}{d} \right)^{\beta_2 - \beta_1} \gg 1 \quad ((\xi_1/d)^{r\xi_3} \gg 1),$$

$$\ln \ln \frac{1}{1-p(d)} \doteq \ln \theta_1 + \beta_1 \ln d$$

or

$$\ln \ln \frac{1}{1-p(d)} \doteq \xi_3 \ln \xi_2 + r\xi_3 \ln \xi_1 + (1-r)\xi_3 \ln d \quad (4.1.5)$$

A plot of $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ will thus approximate to two straight lines for extreme values of $\ln d$, and curve in the middle as shown in Diagram 4.

Diagram 4

Hence P_4/P_2 consists of a family of such curves in the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plots with the two parameters (θ_2, β_2) (or (ζ_2, ζ_3)) describing the behavior at the upper dose range and the other two parameters (θ_1, β_1) (or (ζ_1, r) in combination with the original parameters) describing the behavior at the lower dose range.

The two terms $\theta_1 d^{\beta_1}$ and $\theta_2 d^{\beta_2}$ in the model P_4 thus add flexibility to the original model allowing for possibly different high low dose behaviors. If the high and low dose behaviors are indeed different, at high doses, the term $\theta_2 d^{\beta_2}$ dominates and behavior is described by the old set of parameters (θ_2, β_2) , while at low doses, the term $\theta_1 d^{\beta_1}$ dominates and behavior is described by the extra set of parameters (θ_1, β_1) allowed for in the extended model. The extension may be interpreted in two ways:

1. the difference between $p(d) = 1 - \exp[-\theta_2 d^{\beta_2}]$ and $p(d) = 1 - \exp[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2})]$ is the attachment of a different tail and represents departure from the old model;
2. $p(d) = 1 - \exp[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2})]$ compared with $p(d) = 1 - \exp[-\theta_1 d^{\beta_1}]$ shows an accelerated response rate at the upper dose range.

By allowing β_1 to be free, a full range of low dose behaviors (tails of the hit models) in a continuous scale is thus included in this model.

Many questions may be asked using this model. The choice of model problem corresponds to the choice of tail here. Since the same choice of behaviors in the upper range is allowed, the difficulty with which the data can differentiate between two different tails is the same as that for differentiating between two subclasses of the model with different tails, and may be examined against β_1 and θ_1 , the indices for the different tails attached. In particular, whether or not the low dose behavior is the same as that in the upper range can be tested through testing $\theta_1 = 0$. The extended model may also be used for estimating the low dose behavior either because the high and low dose behaviors are different or so that the estimation can be relatively free from the model dependency problem.

The design aspect for examining the question 'are we using the right tail?' is investigated in Section 4.2 using this proposed extended model. The possibility of determining which is the right tail and its implication on low dose extrapolation are discussed in Section 4.3. It is hoped that provisions for the different high and low dose behaviors,

the continuous scaling of the tails and the incorporation of all the tails in one model through only two extra parameters θ_1, β_1 may simplify our investigation and serve us as a first step for investigations of more flexible and general models.

4.2. Design for Testing Different Low Dose Behaviors

The design problem so far has been directed towards estimation. In Sections 2.3 and 2.4, we assumed the model

$$P_2 = \{p(d) = 1 - \exp[-\theta d^\beta], 0 \leq d < \infty: 0 < \theta < \infty, 0 < \beta < \infty\}$$

and obtained the locally optimal design for estimating the low dose behavior (extrapolated low dose response rate or safe dose) from P_2 . The design was found to be rather stable with change in underlying dose response function, and involved experimentation at two dose levels with response rates roughly at .1 and .99, and an allocation fraction of approximately .6 of the sample size to the lower dose. The fear that the low dose effect may be different from that at higher doses, and hence depart from the model P_2 , lead us to consider the model

$$\begin{aligned} P_4 &= \{p(d) = 1 - \exp[-(\theta_1 d^{\beta_1} + \theta_2 d^{\beta_2})], 0 \leq d < \infty: \\ &\quad 0 \leq \theta_1 < \infty, 0 < \theta_2 < \infty, 0 < \beta_1 < \beta_2 < \infty\} \\ &= \{p(d) = 1 - \exp[-(\zeta_2 d)^{\zeta_3} (1 + (\zeta_1/d)^r \zeta_3)], 0 \leq d < \infty: \\ &\quad 0 \leq \zeta_1 < \infty, 0 < \zeta_2 < \infty, 0 < \zeta_3 < \infty, 0 < r < 1\} \end{aligned}$$

proposed in Section 4.1. The design problem for testing different low dose behaviors is investigated in this section.

The question we address may be put as follows. To what degree is it possible to decide from data whether they originate from a model belonging to $P_2 \subset P_4$, rather than one belonging to $P_4 \setminus P_2$? (The former class of functions assumes the same high and low dose behaviors and the latter assumes different high and low dose behaviors.)

The testing of P_2 against $P_4 \setminus P_2$ is a difficult problem because the difference between the two models depends on how close β_1 is to β_2 , as well as on how close θ_1 and θ_2 are to zero. The design and feasibility of testing between them will likely depend on these parameters.

So instead, we will attempt an easier problem: that of testing P_2 against a three parameter subset $P_3 \setminus P_2$ of $P_4 \setminus P_2$. Our choice of P_3 is

$$\{p(d) = 1 - \exp[-(\zeta_2 d)^{\zeta_3} (1 + (\zeta_1/d)^{\gamma \zeta_3})], 0 \leq d < \infty; 0 \leq \zeta_1 < \infty, \\ 0 < \zeta_2 < \infty, 0 < \zeta_3 < \infty\}$$

with $0 < \gamma < 1$ known. The idea behind the restriction is to allow for different high and low dose behaviors but with only two types of tails so that the design for, and the possibility of distinguishing the two tails can be investigated more thoroughly. If such can be done easily, then the next step would be to extend and differentiate between P_2 and $P_4 \setminus P_2$. Meanwhile, the study should give better understanding of the more general problem.

We will consider the design problem for distinguishing P_2 from $P_3 \setminus P_2$ in this section, and examine the feasibility of distinguishing P_2 from $P_4 \setminus P_2$ in the next section.

We will in particular investigate a special class of designs which seem 'reasonable.' The design strategy here is to assume a large sample experiment, and augment the locally optimal design for estimation from P_2 by adding one extra point to test between the two tails. Since the two points for estimation are very much on the outskirts of the observable range, we will look for a point in between to test for the tails.

The design strategy described above is intended to serve simultaneously two purposes. There are two possible situations:

Case 1. the high and low dose behaviors are really the same, so P_2 is the model;

Case 2. the high and low dose behaviors are different, so $P_3 \setminus P_2$ is the model.

In Case 1, our design provides two points for optimal estimation of the low dose behavior, and in Case 2, the extra one point provides a warning signal for detection of different high and low dose behaviors. Since the third point and its sample size are free to be allocated, a reasonable design for such may be expected even though the optimal design for testing need not be achievable. Further, since the experiment takes only three points, the samples can be concentrated on these three points, and if evidence really shows that the more general model is needed for estimation, another experiment may be designed specially for that purpose.

With the above design strategy, we will proceed to investigate how we should allocate the third dose and its sample size.

Let $\mathcal{C}_2$ and $\mathcal{C}_3$ be the parameter spaces of models P_2 and P_3 . Hence

$$P_3 = \left\{ p_{\underline{\xi}}(d) = 1 - \exp \left[-(\xi_2 d)^{\xi_3} \left(1 + (\xi_1/d)^{\gamma \xi_3} \right) \right], 0 \leq d < \infty: \right.$$

$$\underline{\xi} = \left(\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \end{array} \right) \in \textcircled{\xi}_3 \left. \right\},$$

with $0 < \gamma < 1$ and parameter space

$$\textcircled{\xi}_3 = \left\{ \left(\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \end{array} \right) : 0 \leq \xi_1 < \infty, 0 < \xi_2 < \infty, 0 < \xi_3 < \infty \right\},$$

$$P_2 = \left\{ p_{\underline{\xi}}(d) = 1 - \exp[-(\xi_2 d)^{\xi_3}], 0 \leq d < \infty: \underline{\xi} = \left(\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \end{array} \right) \in \textcircled{\xi}_2 \right\}$$

with parameter space

$$\textcircled{\xi}_2 = \left\{ \left(\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \end{array} \right) : \xi_1 = 0, 0 < \xi_2 < \infty, 0 < \xi_3 < \infty \right\}.$$

Suppose we use P_3 as our model and $\underline{\xi}^0$ is the true value of the parameter. We shall take $x^{11}(\underline{\xi}^0)$ as the design objective because of the following justification. Let the experiment be performed at three doses $b_1 \leq d_1 < d_3 < d_2 \leq b_2$ with sample size $n_i = \rho_i N$ for dose d_i , $i = 1, 2, 3$. If $P_3 \setminus P_2$ holds, then $\underline{\xi}^0$ is an interior point of the

parameter space $\textcircled{\xi}_3 \setminus \textcircled{\xi}_2$, and the other conditions in the result from [5] stated in Section 2.2 hold. Thus the maximum likelihood estimator $\hat{\xi}_1$ of ξ_1 is such that

$$\frac{\hat{\xi}_1 - \xi_1^0}{\sqrt{\frac{x^{11}(\xi^0)}{N}}} \sim \mathcal{N}(0,1)$$

for large N , and if we test $H_0: \xi_1 = 0$ versus $H_1: \xi_1 > 0$ by the criterion: reject H_0 iff $\hat{\xi}_1 > c$ for a fixed $c > 0$,

$$\begin{aligned} \Pr(\text{reject } H_0 | H_1) &= \Pr\left(\frac{\hat{\xi}_1 - \xi_1^0}{\sqrt{x^{11}(\xi^0)/N}} > \frac{c - \xi_1^0}{\sqrt{x^{11}(\xi^0)/N}} \mid \xi^0 \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2\right) \\ &= 1 - \Phi\left(\frac{c - \xi_1^0}{\sqrt{x^{11}(\xi^0)/N}}\right), \quad \xi^0 \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2. \end{aligned}$$

Design to maximize this probability is then equivalent to design to minimize $x^{11}(\xi^0)$.

We will now derive the expression of $x^{11}(\xi)$ for $\xi \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2$.

Let d_1, d_2, d_3 be the three experimental doses with $n_i = \rho_i N$ subjects exposed to dose d_i , $i = 1, 2, 3$. Let

$$q_{\xi}(d) = \ln(1 - p_{\xi}(d))^{-1}$$

$$\frac{\partial p_{\xi}(d)}{\partial \xi_i} = \exp(-q_{\xi}(d)) \frac{\partial q_{\xi}(d)}{\partial \xi_i}, \quad i = 1, 2, 3.$$

Define

$$Q = \left[\frac{\partial q_{\underline{\zeta}}(d_j)}{\partial \zeta_i} \right]_{i,j=1,2,3}$$

and let Q_{1i} be the minor of the $(1,i)$ element of Q . To simplify the calculation for $x^{11}(\underline{\zeta})$, we will first rewrite c_{ℓ} , $\ell = 1,2,3$ in (2.2.4) in terms of Q and Q_{1i} , $i = 1,2,3$. From (2.2.1),

$$B = Q \begin{bmatrix} \exp[-q_{\underline{\zeta}}(d_1)] & 0 & 0 \\ 0 & \exp[-q_{\underline{\zeta}}(d_2)] & 0 \\ 0 & 0 & \exp[-q_{\underline{\zeta}}(d_3)] \end{bmatrix},$$

hence

$$\|B\| = \exp[-(q_{\underline{\zeta}}(d_1) + q_{\underline{\zeta}}(d_2) + q_{\underline{\zeta}}(d_3))] \|Q\|,$$

$$B_{1i} = \frac{\exp[-(q_{\underline{\zeta}}(d_1) + q_{\underline{\zeta}}(d_2) + q_{\underline{\zeta}}(d_3))]}{\exp(-q_{\underline{\zeta}}(d_i))} Q_{1i}, \quad i = 1,2,3,$$

and from (2.2.4),

$$c_{\ell} = \sqrt{\frac{p_{\underline{\zeta}}(d_{\ell})}{1 - p_{\underline{\zeta}}(d_{\ell})}} \left| \frac{Q_{1\ell}}{\|Q\|} \right|, \quad \ell = 1,2,3. \quad (4.2.1)$$

The expressions for Q_{11} , Q_{12} , Q_{13} and $\|Q\|$ are found by direct calculations:

$$Q = \begin{bmatrix} \frac{r\zeta_3}{\zeta_1} [q_{\underline{1}}(d_1) - (\zeta_2 d_1)^{\zeta_3}] & \frac{r\zeta_3}{\zeta_1} [q_{\underline{1}}(d_2) - (\zeta_2 d_2)^{\zeta_3}] & \frac{r\zeta_3}{\zeta_1} [q_{\underline{1}}(d_3) - (\zeta_2 d_3)^{\zeta_3}] \\ \frac{\zeta_3}{\zeta_2} q_{\underline{1}}(d_1) & \frac{\zeta_3}{\zeta_2} q_{\underline{1}}(d_2) & \frac{\zeta_3}{\zeta_2} q_{\underline{1}}(d_3) \\ q_{\underline{1}}(d_1) \ln(\zeta_2 d_1) & q_{\underline{1}}(d_2) \ln(\zeta_2 d_2) & q_{\underline{1}}(d_3) \ln(\zeta_2 d_3) \\ + r \ln\left(\frac{\zeta_1}{d_1}\right) [q_{\underline{1}}(d_1) - (\zeta_2 d_1)^{\zeta_3}] & + r \ln\left(\frac{\zeta_1}{d_2}\right) [q_{\underline{1}}(d_2) - (\zeta_2 d_2)^{\zeta_3}] & + r \ln\left(\frac{\zeta_1}{d_3}\right) [q_{\underline{1}}(d_3) - (\zeta_2 d_3)^{\zeta_3}] \end{bmatrix}$$

Let

$$Z_1 = \left[1 + \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right] \left[1 + \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right] [\ln \xi_2 d_3 - \ln \xi_2 d_2] + r \left(\ln \frac{\xi_1}{d_3} \right) \left[1 + \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right]$$

$$\left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} - r \left(\ln \frac{\xi_1}{d_2} \right) \left[1 + \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right] \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}}$$

$$Z_2 = \left[1 + \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right] \left[1 + \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right] [\ln \xi_2 d_1 - \ln \xi_2 d_3] + r \left(\ln \frac{\xi_1}{d_1} \right) \left[1 + \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right]$$

$$\left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} - r \left(\ln \frac{\xi_1}{d_3} \right) \left[1 + \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right] \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}}$$

$$Z_3 = \left[1 + \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right] \left[1 + \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right] [\ln \xi_2 d_2 - \ln \xi_2 d_1] + r \left(\ln \frac{\xi_1}{d_2} \right) \left[1 + \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right]$$

$$\left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} - r \left(\ln \frac{\xi_1}{d_1} \right) \left[1 + \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right] \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}}$$

$$\begin{aligned} Z = & \left[\left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} - \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right] \left[\ln \xi_2 d_1 + \left(\ln \xi_2 d_1 \right) \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} + r \left(\ln \frac{\xi_1}{d_1} \right) \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right] \\ & + \left[\left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} - \left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} \right] \left[\ln \xi_2 d_2 + \left(\ln \xi_2 d_2 \right) \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} + r \left(\ln \frac{\xi_1}{d_2} \right) \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right] \\ & + \left[\left(\frac{\xi_1}{d_1} \right)^{r_{\xi_3}} - \left(\frac{\xi_1}{d_2} \right)^{r_{\xi_3}} \right] \left[\ln \xi_2 d_3 + \left(\ln \xi_2 d_3 \right) \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} + r \left(\ln \frac{\xi_1}{d_3} \right) \left(\frac{\xi_1}{d_3} \right)^{r_{\xi_3}} \right] \end{aligned}$$

$$\therefore Q_{11} = \frac{\xi_3}{\xi_2} (\xi_2^2 d_2 d_3)^{\xi_3} z_1 \quad (4.2.2)$$

$$-Q_{12} = \frac{\xi_3}{\xi_2} (\xi_2^2 d_1 d_3)^{\xi_3} z_2 \quad (4.2.3)$$

$$Q_{13} = \frac{\xi_3}{\xi_2} (\xi_2^2 d_1 d_2)^{\xi_3} z_3 \quad (4.2.4)$$

after some simplification,

$$\|Q\| = \frac{r \xi_3^2}{\xi_1 \xi_2} (\xi_2^3 d_1 d_2 d_3)^{\xi_3} z \quad (4.2.5)$$

Let

$$f = \frac{n_1}{n_1 + n_2}, \quad \epsilon = \frac{n_3}{n_1 + n_2},$$

the allocation fractions are

$$\rho_1 = \frac{f}{1 + \epsilon}, \quad \rho_2 = \frac{1 - f}{1 + \epsilon}, \quad \rho_3 = \frac{\epsilon}{1 + \epsilon} \quad (4.2.6)$$

Hence if $\underline{\xi} \in \mathbb{C}_3 \setminus \mathbb{C}_2$ obtains, the expression for $x^{11}(\underline{\xi})$ from (2.2.3) and (4.2.6) is:

$$x^{11}(\underline{\xi}) = \left(\frac{c_1^2}{f} + \frac{c_2^2}{1 - f} + \frac{c_3^2}{\epsilon} \right) (1 + \epsilon)$$

where c_ℓ , $\ell = 1, 2, 3$ are given by (4.2.1), $Q_{1\ell}$, $\ell = 1, 2, 3$ and $\|Q\|$ are given by (4.2.2) to (4.2.5).

The expression of $x^{11}(\underline{\xi})$ involves d_1, d_2, d_3, f and ϵ , but since d_1, d_2 and f are fixed for roughly optimal estimation of

$p^0(d_0)$ (or $d^0(p_0)$) from P_2 , d_3 and ϵ are the only free parameters to be decided upon.

If we set the derivative of $x^{11}(\underline{\zeta})$ with respect to ϵ to zero, we have

$$\frac{c_1^2}{f} + \frac{c_2^2}{1-f} + \frac{c_3^2}{\epsilon} - (1+\epsilon) \frac{c_3^2}{\epsilon^2} = 0$$

hence

$$\epsilon = \frac{c_3}{\sqrt{\frac{c_1^2}{f} + \frac{c_2^2}{1-f}}} \quad (4.2.7)$$

$$1 + \epsilon = \frac{c_3 + \sqrt{\frac{c_1^2}{f} + \frac{c_2^2}{1-f}}}{\sqrt{\frac{c_1^2}{f} + \frac{c_2^2}{1-f}}}$$

With this allocation fraction,

$$x^{11}(\underline{\zeta}) = \left(c_3 + \sqrt{\frac{c_1^2}{f} + \frac{c_2^2}{1-f}} \right)^2$$

$$= \frac{\zeta_1^2}{r^2 \zeta_3^2 z^2} \left\{ \left[\frac{p_{\underline{\zeta}}(d_1)}{f(1-p_{\underline{\zeta}}(d_1))} \frac{z_1^2}{(\zeta_2 d_1)^{2\zeta_3}} + \frac{p_{\underline{\zeta}}(d_2)}{(1-f)(1-p_{\underline{\zeta}}(d_2))} \frac{z_2^2}{(\zeta_2 d_2)^{2\zeta_3}} \right]^{1/2} + \left[\frac{p_{\underline{\zeta}}(d_3)}{(1-p_{\underline{\zeta}}(d_3))} \frac{z_3^2}{(\zeta_2 d_3)^{2\zeta_3}} \right]^{1/2} \right\}^2$$

Suppose $\underline{\zeta}$ is known to be close to $\underline{\zeta}^0 \in \mathbb{C}_3 \setminus \mathbb{C}_2$ so that $x^{11}(\underline{\zeta}^0)$ is to be minimized over d_3 . Since d_1 and d_2 are already close to the boundaries of the observable region, we will look for a value of d_3 in between d_1 and d_2 for the minimization. (Such

a minimum between d_1 and d_2 exists since $x^{11}(\xi^0) \geq 0$ for all $0 \leq d_3 < \infty$, and as $d_3 \rightarrow d_1$ or d_2 , $Z|_{\xi^0} \rightarrow 0$ and $x^{11}(\xi^0) \rightarrow \infty$. For easier investigation, we will let $\frac{\xi = \xi^0}{d_3} = d_1^{1/2-\lambda} d_2^{1/2+\lambda}$, $-1/2 < \lambda < 1/2$ and consider $x^{11}(\xi^0)$ also as a function of λ . When regarded as a function of λ , we will seek a value of λ between $-1/2$ and $1/2$ to minimize $x^{11}(\xi^0)$. The minimizing value of λ is then denoted by $\lambda(\xi^0)$ and that of d_3 by $d_3(\xi^0)$; i.e. $d_3(\xi^0) = d_1^{1/2-\lambda(\xi^0)} d_2^{1/2+\lambda(\xi^0)}$.

Hence from (4.2.7) the optimal allocation fraction $\epsilon(\xi^0)$

$$= \left[\frac{\frac{p_{\xi}(d_3)}{(1-p_{\xi}(d_3))} \cdot \frac{z_3^2}{(\xi_2 d_3)^{2\xi_3}}}{\frac{p_{\xi}(d_1)}{f(1-p_{\xi}(d_1))} \cdot \frac{z_1^2}{(\xi_2 d_1)^{2\xi_3}} + \frac{p_{\xi}(d_2)}{(1-f)(1-p_{\xi}(d_2))} \cdot \frac{z_2^2}{(\xi_2 d_2)^{2\xi_3}}} \right]^{1/2} \Big|_{\substack{\xi = \xi^0 \\ d_3 = d_3(\xi^0)}}$$

may also be regarded as a function of $d_3(\xi^0)$ or as a function of $\lambda(\xi^0)$.

Hence the final design augments the two point design (which allocates d_1, d_2 at $p^0(d_1) = 1, p^0(d_2) = .99$ and

$$f = \left[1 + \sqrt{\frac{.99(.9)}{.01(.1)} \frac{\ln(1-.1)^{-1}}{\ln(1-.99)^{-1}} \left| \frac{\ln \ln(1-.1)^{-1} - \ln \ln(1-p^0(d_0))^{-1}}{\ln \ln(1-.99)^{-1} - \ln \ln(1-p^0(d_0))^{-1}} \right| \right]^{-1}$$

to optimally estimate the low dose response rate (or safe dose) from P_2 with a third point located at $d_3 = d_3(\xi^0)$ and with an allocation fraction $\epsilon = \epsilon(\xi^0)$ when $\xi^0 \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2$ obtains.

Note that if P_3 is the model, the exact location of τ_1^0 and τ_2^0 depend on $\xi^0 \in \mathbb{C}_3$. The allocation of d_1 and d_2 to τ_1^0 and τ_2^0 depend on the information from the preliminary experiment for pinning down these two points in the observable region and the allocation can be made accurate without the knowledge of $\xi^0 \in \mathbb{C}_3$, or whether $\xi^0 \in \mathbb{C}_2$ or $\xi^0 \in \mathbb{C}_3 \setminus \mathbb{C}_2$. The exact location of f on the other hand does not depend on $\xi^0 \in \mathbb{C}_3$. It depends on the unknown $p^0(d_0)$ and can be fixed only at an approximate optimal value.

In order to explore how $d_3(\xi^0)$ (or $\lambda(\xi^0)$) and $\epsilon(\xi^0)$ depend on the parameter $\xi^0 \in \mathbb{C}_3 \setminus \mathbb{C}_2$, we will first establish another parametrization of the model $P_3 \setminus P_2$.

Let

$$\ell_1 = \ln(1 - .1)^{-1}, \quad \ell_2 = \ln(1 - .99)^{-1}$$

Given

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \in \mathbb{C}_3 \setminus \mathbb{C}_2,$$

let

$$\tau_1 \text{ be defined by } p_\xi(\tau_1) = .1,$$

$$\tau_2 \text{ be defined by } p_\xi(\tau_2) = .99,$$

$$\tau_3 = \xi_3.$$

Since $p_\xi(d)$ is strictly increasing from 0 to 1 as d increases from 0 to ∞ ,

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$$

is well defined, $0 < \tau_1 < \tau_2 < \infty$, $0 < \tau_3 < \infty$, and is related to $\underline{\xi}$ by the equations:

$$\tau_3 = \xi_3$$

$$\begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix} = \begin{bmatrix} \xi_3 & (1-\gamma)\xi_3 \\ \tau_1 & \tau_1 \\ \xi_3 & (1-\gamma)\xi_3 \\ \tau_2 & \tau_2 \end{bmatrix} \begin{pmatrix} \xi_3 \\ \xi_2 \\ \xi_3 \\ \xi_1 \end{pmatrix}$$

i.e. $\tau_3 = \xi_3$

$$\begin{pmatrix} \xi_3 \\ \xi_2 \\ \xi_3 \\ \xi_1 \end{pmatrix} = \frac{1}{(\tau_1 \tau_2)^{\tau_3} - \tau_2^{\tau_3} \tau_1^{\tau_3}} \begin{bmatrix} (1-\gamma)\tau_3 & (1-\gamma)\tau_3 \\ \tau_2 & -\tau_1 \\ -\tau_2 & \tau_1 \end{bmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix} \quad (4.2.8)$$

For $\underline{\xi} \in \mathbb{C}_3 \setminus \mathbb{C}_2$,

$$0 < \xi_1 < \infty, 0 < \xi_2 < \infty \implies \left(\frac{\ell_1}{\ell_2}\right)^{1/(1-\gamma)} < \left(\frac{\tau_1}{\tau_2}\right)^{\tau_3} < \frac{\ell_1}{\ell_2}.$$

The parameter space $\mathbb{C}_3 \setminus \mathbb{C}_2$ thus transforms to the space

$$\mathbb{T} = \left\{ \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} : 0 < \tau_i < \infty, i = 1, 2, 3, \left(\frac{\ell_1}{\ell_2}\right)^{1/(1-\gamma)} < \left(\frac{\tau_1}{\tau_2}\right)^{\tau_3} < \frac{\ell_1}{\ell_2} \right\}.$$

Given $\underline{\xi} \in \mathbb{C}_3 \setminus \mathbb{C}_2$, there corresponds $\underline{\tau} \in \mathbb{T}$, and conversely, given $\underline{\tau} \in \mathbb{T}$, (4.2.8) defines $\underline{\xi} \in \mathbb{C}_3 \setminus \mathbb{C}_2$ corresponding to $\underline{\tau}$. The model $P_3 \setminus P_2$ can hence be written as,

$$\begin{aligned}
P_3 \setminus P_2 &= \left\{ p_{\underline{\xi}}(d) = 1 - \exp \left[-(\xi_2 d)^{\xi_3} \left(1 + \left(\frac{\xi_1}{d} \right)^{\tau \xi_3} \right) \right], 0 \leq d < \infty; \underline{\xi} \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2 \right\} \\
&= \left\{ p_{\underline{\tau}}(d) = 1 - \exp \left\{ - \left[d^{\tau_3} (\tau_2^{(1-\tau)\tau_3} \tau_1^{(1-\tau)\tau_3} \right) \right. \right. \\
&\quad \left. \left. + d^{(1-\tau)\tau_3} (\tau_2^{-\tau_3} \tau_1^{\tau_3}) \right] / \left[\tau_1^{\tau_3} \tau_2^{(1-\tau)\tau_3} \tau_3^{-\tau_3} \tau_1^{(1-\tau)\tau_3} \right] \right\}, \\
&\quad 0 \leq d < \infty; \underline{\tau} \in \textcircled{\tau} \right\}.
\end{aligned}$$

The one-one correspondence between $\underline{\xi}$ and $\underline{\tau}$ implies that each dose response function $P_3 \setminus P_2$ can be referred by a $\underline{\xi}$ value or by a $\underline{\tau}$ value. If $\underline{\xi}^0, \underline{\tau}^0$ are the corresponding true parameter values, $x^{11}(\underline{\xi}^0)$, $d_3(\underline{\xi}^0)$, $\lambda(\underline{\xi}^0)$ and $\epsilon(\underline{\xi}^0)$ may also be regarded as functions of $\underline{\tau}^0$ and we will use $x^{11}(\underline{\tau}^0)$, $d_3(\underline{\tau}^0)$, $\lambda(\underline{\tau}^0)$, $\epsilon(\underline{\tau}^0)$ to denote the corresponding quantities when we are using the $\underline{\tau}$ parametrization. Note also that, if $\underline{\xi}^0 \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2$ obtains, then the two doses for the original two point design should be allocated to $d_1 = \tau_1^0$, $d_2 = \tau_2^0$.

Instead of examining how $\lambda(\underline{\xi}^0)$ and $\epsilon(\underline{\xi}^0)$ depend on $\underline{\xi}^0 \in \textcircled{\xi}_3 \setminus \textcircled{\xi}_2$, we will examine how $\lambda(\underline{\tau}^0)$ and $\epsilon(\underline{\tau}^0)$ depend on $\underline{\tau}^0 \in \textcircled{\tau}$. We will show that if $\underline{\tau}^{0(1)}$ is related to $\underline{\tau}^{0(2)}$ in certain ways then for every $-1/2 < \lambda < 1/2$,

$$x^{11}(\underline{\tau}^{0(1)}) = m x^{11}(\underline{\tau}^{0(2)})$$

and

$$\epsilon(\underline{\tau}^{0(1)}) = \epsilon(\underline{\tau}^{0(2)})$$

where m does not depend on λ . Therefore when $\underline{\tau}^{0(1)}$ or $\underline{\tau}^{0(2)}$ obtains,

$$\lambda(\underline{\tau}^{0(1)}) = \lambda(\underline{\tau}^{0(2)})$$

$$\epsilon(\underline{\tau}^{0(1)}) = \epsilon(\underline{\tau}^{0(2)})$$

i.e. d_3 in relation to d_1, d_2 and the allocation fraction would be the same. Hence we need investigate $\lambda(\underline{\tau}^0), \epsilon(\underline{\tau}^0)$ for unrelated $\underline{\tau}^0$'s in $\odot$.

The proof comes in two parts:

1. Let

$$\underline{\tau}^{(1)} = \begin{pmatrix} \tau_1^{(1)} \\ \tau_2^{(1)} \\ \tau_3^{(1)} \end{pmatrix} \quad \text{and} \quad \underline{\tau}^{(2)} = \begin{pmatrix} \tau_1^{(2)} \\ \tau_2^{(2)} \\ \tau_3^{(2)} \end{pmatrix}$$

with $\tau_3^{(1)} = \tau_3^{(2)}$. We shall show that if

$$\frac{\tau_1^{(1)}}{\tau_2^{(1)}} = \frac{\tau_1^{(2)}}{\tau_2^{(2)}}$$

then $\lambda(\underline{\tau}^{(1)}) = \lambda(\underline{\tau}^{(2)})$ and $\epsilon(\underline{\tau}^{(1)}) = \epsilon(\underline{\tau}^{(2)})$.

Proof. (1) Let $\tau_3^{(1)} = \tau_3^{(2)} = \tau_3, \tau_1^{(2)} = c\tau_1^{(1)}$ for some $c > 0$ then $\tau_2^{(2)} = c\tau_2^{(1)}$, i.e.

$$\underline{\tau}^{(1)} = \begin{pmatrix} \tau_1^{(1)} \\ \tau_2^{(1)} \\ \tau_3 \end{pmatrix}, \quad \underline{\tau}^{(2)} = \begin{pmatrix} c\tau_1^{(1)} \\ c\tau_2^{(1)} \\ \tau_3 \end{pmatrix}$$

(2)

Let

$$\underline{\zeta}^{(1)} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \tau_3 \end{pmatrix}$$

correspond to $\underline{\tau}^{(1)}$, therefore from (4.2.8) they are related by

$$\begin{pmatrix} \alpha_2^{\tau_3} \\ \alpha_2^{\tau_3} & \alpha_1^{\tau_3} \end{pmatrix} = \begin{bmatrix} (1)\tau_3 & (1)(1-\gamma)\tau_3 & (1)\tau_3 & (1)(1-\gamma)\tau_3 \\ \tau_1 & \tau_2 & -\tau_2 & \tau_1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} (1)(1-\gamma)\tau_3 & (1)(1-\gamma)\tau_3 \\ \tau_2 & -\tau_1 \\ -(1)\tau & (1)\tau_3 \\ -\tau_2 & \tau_1 \end{bmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} (\alpha_2/c)^{\tau_3} \\ (\alpha_2/c)^{\tau_3} & (\alpha_1/c)^{\tau_3} \end{pmatrix}$$

$$= \begin{bmatrix} (c\tau_1^{(1)})^{\tau_3} & (c\tau_2^{(1)})^{(1-\gamma)\tau_3} & -(c\tau_2^{(1)})^{\tau_3} & (c\tau_1^{(1)})^{(1-\gamma)\tau_3} \end{bmatrix}^{-1} \cdot \begin{bmatrix} (c\tau_2^{(1)})^{(1-\gamma)\tau_3} & -(c\tau_1^{(1)})^{(1-\gamma)\tau_3} \\ -(c\tau_2^{(1)})^{\tau_3} & (c\tau_1^{(1)})^{\tau_3} \end{bmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}$$

Hence if $\underline{\zeta}^{(2)}$ is the vector corresponding to $\underline{\tau}^{(2)}$, then

$$\underline{\zeta}^{(2)} = \begin{pmatrix} c\alpha_1 \\ \frac{\alpha_2}{c} \\ \tau_3 \end{pmatrix}$$

(3)

Recall that if $\underline{\zeta} = \underline{\zeta}^{(1)}$ obtains then $d_1 = \tau_1^{(1)}$, $d_2 = \tau_2^{(1)}$
and if $\underline{\zeta} = \underline{\zeta}^{(2)}$ obtains then $d_1 = \tau_1^{(2)}$, $d_2 = \tau_2^{(2)}$, therefore

$$\left. \frac{\zeta_1}{d_i} \right|_{\underline{\zeta}=\underline{\zeta}^{(1)}} = \frac{\alpha_1}{\tau_i^{(1)}} = \frac{\alpha_1 c}{c\tau_i^{(1)}} = \left. \frac{\zeta_1}{d_i} \right|_{\underline{\zeta}=\underline{\zeta}^{(2)}}$$

$$\left. \zeta_2 d_i \right|_{\underline{\zeta}=\underline{\zeta}^{(1)}} = \alpha_2 \tau_i^{(1)} = (\alpha_2/c) (c\tau_i^{(1)}) = \left. \zeta_2 d_i \right|_{\underline{\zeta}=\underline{\zeta}^{(2)}}$$

for $i = 1, 2$.

Also

$$d_3 = d_1^{1/2} d_2^{1/2} (d_2/d_1)^\lambda$$

$$\begin{aligned} \left. \frac{\zeta_1}{d_3} \right|_{\underline{\zeta}=\underline{\zeta}^{(1)}} &= \frac{\alpha_1}{\tau_1^{(1)1/2} \tau_2^{(1)1/2}} \left(\frac{\tau_1^{(1)}}{\tau_2^{(1)}} \right)^\lambda \\ &= \frac{\alpha_1 c}{(c\tau_1^{(1)})^{1/2} (c\tau_2^{(1)})^{1/2}} \left(\frac{c\tau_1^{(1)}}{c\tau_2^{(1)}} \right)^\lambda \\ &= \left. \frac{\zeta_1}{d_3} \right|_{\underline{\zeta}=\underline{\zeta}^{(2)}} \end{aligned}$$

$$\begin{aligned}
\zeta_2 d_3 \Big|_{\underline{\zeta}=\underline{\zeta}(1)} &= \alpha_2 \tau_1^{(1)1/2} \tau_2^{(1)1/2} \left(\frac{\tau_2^{(1)}}{\tau_1^{(1)}} \right)^\lambda \\
&= \left(\frac{\alpha_2}{c} \right) (c\tau_1^{(1)})^{1/2} (c\tau_2^{(1)})^{1/2} \left(\frac{c\tau_2^{(1)}}{c\tau_1^{(1)}} \right)^\lambda \\
&= \zeta_2 d_3 \Big|_{\underline{\zeta}=\underline{\zeta}(2)}
\end{aligned}$$

Hence $p_{\underline{\zeta}}(d_i) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = p_{\underline{\zeta}}(d_i) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$, $i = 1, 2$ and for all

$$-1/2 < \lambda < 1/2 ,$$

$$z_i \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = z_i \Big|_{\underline{\zeta}=\underline{\zeta}(2)} , \quad i = 1, 2, 3$$

$$z \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = z \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$p_{\underline{\zeta}}(d_3) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = p_{\underline{\zeta}}(d_3) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$\Rightarrow x^{11}(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = c^{-2} x^{11}(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$\epsilon(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = \epsilon(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$\text{hence } \lambda(\underline{\zeta}^{(1)}) = \lambda(\underline{\zeta}^{(2)}) , \quad \epsilon(\underline{\zeta}^{(1)}) = \epsilon(\underline{\zeta}^{(2)})$$

$$\text{i.e. } \lambda(\underline{\tau}^{(1)}) = \lambda(\underline{\tau}^{(2)}) , \quad \epsilon(\underline{\tau}^{(1)}) = \epsilon(\underline{\tau}^{(2)}) .$$

2. Let

$$\underline{\tau}^{(1)} = \begin{pmatrix} \tau_1^{(1)} \\ \tau_2^{(1)} \\ \tau_3^{(1)} \end{pmatrix} \quad \text{and} \quad \underline{\tau}^{(2)} = \begin{pmatrix} \tau_1^{(2)} \\ \tau_2^{(2)} \\ \tau_3^{(2)} \end{pmatrix}$$

with $\tau_2^{(1)} = \tau_2^{(2)} = 1$. We shall show that if $\tau_1^{(1)} = \tau_1^{(2)} \tau_3^{(2)} / \tau_3^{(1)}$ then $\lambda(\underline{\tau}^{(1)}) = \lambda(\underline{\tau}^{(2)})$, $\epsilon(\underline{\tau}^{(1)}) = \epsilon(\underline{\tau}^{(2)})$.

Proof. (1) $\tau_2^{(1)} = \tau_2^{(2)} = 1$, $\tau_3^{(2)} = c \tau_3^{(1)}$, $c > 0$ then

$$\tau_1^{(2)} = \tau_1^{(1)} / c, \text{ i.e.}$$

$$\underline{\tau}^{(1)} = \begin{pmatrix} \tau_1^{(1)} \\ 1 \\ \tau_3^{(1)} \end{pmatrix}, \quad \underline{\tau}^{(2)} = \begin{pmatrix} \tau_1^{(1)} / c \\ 1 \\ c \tau_3^{(1)} \end{pmatrix}$$

(2)

Let

$$\underline{\xi}^{(1)} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \tau_3^{(1)} \end{pmatrix}$$

corresponds to $\underline{\tau}^{(1)}$, therefore $\underline{\xi}^{(1)}$, $\underline{\tau}^{(1)}$ are related by

$$\begin{pmatrix} \alpha_2^{\tau_3^{(1)}} \\ \alpha_2^{\tau_3^{(1)}} & \alpha_1^{\tau_3^{(1)}} \end{pmatrix}$$

$$= \begin{bmatrix} (1)\tau_3^{(1)} & (1)(1-r)\tau_3^{(1)} \\ \tau_1 & -\tau_1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\tau_1^{(1)(1-r)\tau_3^{(1)}} \\ -1 & \tau_1^{(1)\tau_3^{(1)}} \end{bmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} (\alpha_2^{1/c})^{c\tau_3^{(1)}} \\ (\alpha_2^{1/c})^{c\tau_3^{(1)}} & (\alpha_1^{1/c})^{c\tau_3^{(1)}} \end{pmatrix}$$

$$= \begin{bmatrix} (1)1/c \quad c\tau_3^{(1)} & (1)1/c \quad c(1-r)\tau_3^{(1)} \\ (\tau_1) & -(\tau_1) \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\tau_1^{(1)(1/c)c(1-r)\tau_3^{(1)}} \\ -1 & \tau_1^{(1)(1/c)c\tau_3^{(1)}} \end{bmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}$$

Hence if $\underline{\xi}^{(2)}$ corresponds to $\underline{\tau}^{(2)}$ then

$$\underline{\xi}^{(2)} = \begin{pmatrix} \alpha_1^{1/c} \\ \alpha_2^{1/c} \\ c\tau_3^{(1)} \end{pmatrix}$$

$$(3) \quad \left(\frac{\xi_1}{d_1} \right)^{\xi_3} \Big|_{\xi=\xi(1)} = \left(\frac{\alpha_1}{\tau_1^{(1)}} \right)^{\tau_3^{(1)}} = \left(\frac{\alpha_1^{1/c}}{\tau_1^{(1)1/c}} \right)^{c\tau_3^{(1)}} = \left(\frac{\xi_1}{d_1} \right)^{\xi_3} \Big|_{\xi=\xi(2)}.$$

$$\left(\frac{\xi_1}{d_2} \right)^{\xi_3} \Big|_{\xi=\xi(1)} = \left(\frac{\alpha_1}{1} \right)^{\tau_3^{(1)}} = \left(\frac{\alpha_1^{1/c}}{1} \right)^{c\tau_3^{(1)}} = \left(\frac{\xi_1}{d_2} \right)^{\xi_3} \Big|_{\xi=\xi(2)}$$

$$(\xi_2 d_1)^{\xi_3} \Big|_{\xi=\xi(1)} = (\alpha_2 \tau_1^{(1)})^{\tau_3^{(1)}} = (\alpha_2^{1/c} \tau_1^{(1)1/c})^{c\tau_3^{(1)}} = (\xi_2 d_1)^{\xi_3} \Big|_{\xi=\xi(2)}$$

$$(\xi_2 d_2)^{\xi_3} \Big|_{\xi=\xi(1)} = \alpha_2^{\tau_3^{(1)}} = (\alpha_2^{1/c})^{c\tau_3^{(1)}} = (\xi_2 d_2)^{\xi_3} \Big|_{\xi=\xi(2)}$$

Also,

$$\begin{aligned} \left(\frac{\xi_1}{d_3} \right)^{\xi_3} \Big|_{\xi=\xi(1)} &= \left(\frac{\alpha_1}{\tau_1^{(1)1/2}} \right)^{\tau_3^{(1)}} (\tau_1^{(1)})^{\tau_3^{(1)}\lambda} \\ &= \left(\frac{\alpha_1^{1/c}}{\tau_1^{(1)1/2 \cdot 1/c}} \right)^{c\tau_3^{(1)}} (\tau_1^{(1)1/c})^{c\tau_3^{(1)}\lambda} \\ &= \left(\frac{\xi_1}{d_3} \right)^{\xi_3} \Big|_{\xi=\xi(2)} \end{aligned}$$

$$\begin{aligned} (\xi_2 d_3)^{\xi_3} \Big|_{\xi=\xi(1)} &= (\alpha_2 \tau_1^{(1)1/2})^{\tau_3^{(1)}} (1/\tau_1^{(1)})^{\lambda \tau_3^{(1)}} \\ &= (\alpha_2^{1/c} \tau_1^{(1)1/(2c)})^{c\tau_3^{(1)}} (1/\tau_1^{(1)1/c})^{\lambda c\tau_3^{(1)}} \\ &= (\xi_2 d_3)^{\xi_3} \Big|_{\xi=\xi(2)} \end{aligned}$$

Hence $P_{\underline{\zeta}}(d_i) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = P_{\underline{\zeta}}(d_i) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$, $i = 1, 2$ and for all

$$-1/2 < \lambda < 1/2.$$

$$\zeta_3^{Z_i} \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = \zeta_3^{Z_i} \Big|_{\underline{\zeta}=\underline{\zeta}(2)}, \quad i = 1, 2, 3$$

$$\begin{aligned} Z_i \Big|_{\underline{\zeta}=\underline{\zeta}(1)} &= \zeta_3 \frac{Z_i}{\tau_3^{(1)}} \Big|_{\underline{\zeta}=\underline{\zeta}(1)} \\ &= \zeta_3 \frac{Z_i}{\tau_3^{(1)}} \Big|_{\underline{\zeta}=\underline{\zeta}(2)} = cZ_i \Big|_{\underline{\zeta}=\underline{\zeta}(2)} \end{aligned}$$

$$Z \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = cZ \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$P_{\underline{\zeta}}(d_3) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = P_{\underline{\zeta}}(d_3) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$\Rightarrow x^{11}(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(1)} = \alpha_1^{2(1-1/c)} c x^{11}(\underline{\zeta}) \Big|_{\underline{\zeta}=\underline{\zeta}(2)}$$

$$\epsilon(\underline{\zeta}^{(1)}) = \epsilon(\underline{\zeta}^{(2)})$$

hence

$$\lambda(\underline{\zeta}^{(1)}) = \lambda(\underline{\zeta}^{(2)}), \quad \epsilon(\underline{\zeta}^{(1)}) = \epsilon(\underline{\zeta}^{(2)})$$

$$\text{i.e.} \quad \lambda(\underline{\tau}^{(1)}) = \lambda(\underline{\tau}^{(2)}), \quad \epsilon(\underline{\tau}^{(1)}) = \epsilon(\underline{\tau}^{(2)})$$

The two parts imply that if

$$\underline{\tau}^{(1)} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}, \quad \underline{\tau}^{(2)} = \begin{pmatrix} (\tau_1/\tau_2)^{\tau_3} \\ 1 \\ 1 \end{pmatrix} \in \textcircled{\tau},$$

then

$$\lambda(\underline{\tau}^{(1)}) = \lambda \left(\begin{pmatrix} \tau_1/\tau_2 \\ 1 \\ \tau_3 \end{pmatrix} \right) = \lambda(\underline{\tau}^{(2)})$$

$$\epsilon(\underline{\tau}^{(1)}) = \epsilon \left(\begin{pmatrix} \tau_1/\tau_2 \\ 1 \\ \tau_3 \end{pmatrix} \right) = \epsilon(\underline{\tau}^{(2)})$$

or equivalently, if

$$\underline{\zeta}^{(1)} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}, \quad \underline{\zeta}^{(2)} = \begin{pmatrix} (\zeta_1 \zeta_2)^{\zeta_3} \\ 1 \\ 1 \end{pmatrix} \in \mathbb{C}_3 \setminus \mathbb{C}_2$$

then

$$\lambda(\underline{\zeta}^{(1)}) = \lambda \left(\begin{pmatrix} \zeta_1 \zeta_2 \\ 1 \\ \zeta_3 \end{pmatrix} \right) = \lambda(\underline{\zeta}^{(2)})$$

$$\epsilon(\underline{\zeta}^{(1)}) = \epsilon \left(\begin{pmatrix} \zeta_1 \zeta_2 \\ 1 \\ \zeta_3 \end{pmatrix} \right) = \epsilon(\underline{\zeta}^{(2)})$$

Hence instead of examining $\lambda(\underline{\tau})$, $\epsilon(\underline{\tau})$ for $\underline{\tau} \in \mathbb{C}_3$, we can examine only for

$$\underline{\tau} \in \left\{ \begin{pmatrix} w \\ 1 \\ 1 \end{pmatrix} : \left(\frac{\ell_1}{\ell_2} \right)^{1/(1-r)} < w < \ell_1/\ell_2 \right\},$$

or instead of examining $\lambda(\underline{\zeta})$, $\epsilon(\underline{\zeta})$ for $\underline{\zeta} \in \textcircled{\zeta}_3 \setminus \textcircled{\zeta}_2$, we can examine only for

$$\underline{\zeta} \in \left\{ \begin{pmatrix} w \\ 1 \\ 1 \end{pmatrix} : 0 < w < \infty \right\}$$

Table 23 shows $\lambda(\underline{\tau})$ and $\epsilon(\underline{\tau})$ for

$$\underline{\tau} = \begin{pmatrix} w \\ 1 \\ 1 \end{pmatrix}, \quad \left(\frac{l_1}{l_2} \right)^{1/(1-\gamma)} < w < l_1/l_2,$$

and with $\gamma = .5$. The original two point design has $p^0(d_1) = .1$, $p^0(d_2) = .99$ and $f = .65$.

The results indicate that design for testing between two tails is already a complicated problem. $\lambda(\underline{\tau})$ is positive for all $\underline{\tau} \in \textcircled{\tau}$. For $\underline{\tau} \in \textcircled{\tau}$ with small $(\tau_1/\tau_2)^{\tau_3}$ (referring to Table 23 for cases with small w), $\lambda(\underline{\tau})$ is generally near .45 but rather erratic, for $\underline{\tau} \in \textcircled{\tau}$ with large $(\tau_1/\tau_2)^{\tau_3}$, say $> .004$ (referring to Table 23 for cases with large w), $\lambda(\underline{\tau})$ decreases smoothly from .3 to .1. The third dose should thus be allocated to somewhere close to d_2 for $\underline{\tau}$'s with small $(\tau_1/\tau_2)^{\tau_3}$ and move towards d_1 as the value of $(\tau_1/\tau_2)^{\tau_3}$ increases. $\epsilon(\underline{\tau})$, the ratio of the sample size for the third dose to the total sample size for the first two doses, is also difficult to assign: the numbers are erratic over the entire range of w . Hence for $\underline{\tau} \in \textcircled{\tau}$ with small $(\tau_1/\tau_2)^{\tau_3}$, $\epsilon(\underline{\tau})$ is close to .5 suggesting that the samples should be more or less evenly divided among the three doses; for $\underline{\tau} \in \textcircled{\tau}$ with large $(\tau_1/\tau_2)^{\tau_3}$, $\epsilon(\underline{\tau})$ is generally very small.

TABLE 23

$\lambda(\underline{\tau})$ and $\epsilon(\underline{\tau})$ for $\underline{\tau} = \begin{pmatrix} w \\ 1 \\ 1 \end{pmatrix}$, $\left(\frac{\ell_1}{\ell_2}\right)^{1/(1-\gamma)} < w < \frac{\ell_1}{\ell_2}$, $\gamma = \frac{1}{2}$.

w	$\lambda(\underline{\tau})$	$\epsilon(\underline{\tau})$
5.5×10^{-4}	.4	.55
6×10^{-4}	.37	.52
7×10^{-4}	.4	.51
8×10^{-4}	.45	.53
9×10^{-4}	.48	5.9×10^{-2}
1×10^{-3}	.46	3.2×10^{-2}
2×10^{-3}	.47	9.6×10^{-2}
3×10^{-3}	.32	6.1×10^{-4}
4×10^{-3}	.29	3.3×10^{-3}
5×10^{-3}	.27	1.8×10^{-3}
6×10^{-3}	.25	4.2×10^{-3}
7×10^{-3}	.23	8.5×10^{-3}
8×10^{-3}	.22	5.0×10^{-3}
9×10^{-3}	.21	2.7×10^{-3}
1×10^{-2}	.2	1.5×10^{-3}
1.1×10^{-2}	.19	8.3×10^{-4}
1.2×10^{-2}	.17	5.5×10^{-4}
1.3×10^{-2}	.17	7.3×10^{-3}
1.4×10^{-2}	.16	4.7×10^{-3}
1.5×10^{-2}	.15	5.7×10^{-3}
1.6×10^{-2}	.14	7.3×10^{-3}
1.7×10^{-2}	.13	9.2×10^{-3}
1.8×10^{-2}	.13	1.1×10^{-2}
1.9×10^{-2}	.12	3.2×10^{-3}
2×10^{-2}	.11	6.0×10^{-3}
2.1×10^{-2}	.1	8.9×10^{-3}
2.2×10^{-2}	.1	1.7×10^{-3}

and hence the total sample size has to be of the order of 10^3 or above to bring the sample size at d_3 to a reasonable number.

4.3. Conclusion

A design for distinguishing P_2 from $P_3 \setminus P_2$ has been examined in Section 4.2. The result indicates some general trends as to how we should allocate the third dose and its sample size; it also shows a heavy dependence of these allocations on the parameters, and the need for a large sample experiment. The design for distinguishing P_2 from $P_4 \setminus P_2$ is expected to be even more complicated. In this section, we will examine the feasibility of distinguishing P_2 from $P_4 \setminus P_2$ and hope to gain some insight into the more general problem of distinguishing between tails.

Recall from Section 4.1 the models P_2 , $P_4 \setminus P_2$, and their $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plots (Diagrams 3 and 4 are reproduced here for convenience).

Diagram 5

Diagram 6

The model P_2 consists of a family of straight lines in the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plots (Diagram 5) and the constant slope of the dose response function in such a plot reflects the assumption that the high and low dose behaviors are the same. The model $P_4 \setminus P_2$, on the other hand, consists of a family of curves in the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plots (Diagram 6), each of which is characterized by a pair of approximating straight lines at extreme values of $\ln d$. The different slopes of the straight lines suggest different high and low dose behaviors.

The model P_4 thus incorporates the dose response functions in P_2 and $P_4 \setminus P_2$ and allows for possibly same or different high low dose behaviors. The departure from P_2 in $P_4 \setminus P_2$ is reflected by the deflection of the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ curve from that of a straight line.

We will first give interpretation to the parameters of $P_4 \setminus P_2$ and then investigate departure from P_2 by examining how well the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ curves of $P_4 \setminus P_2$ approximate to straight lines.

The four parameters of the model $P_4 \setminus P_2$ can be interpreted by the characteristics of the two approximating straight lines. The upper straight line has slope ξ_3 , the lower straight line has slope at a fraction $1-\gamma$ of that of the upper straight line, and the two approximating straight lines intersect at the point $(\ln d, \ln \ln(1-p(d))^{-1}) = (\ln \xi_1, \ln(\xi_1 \xi_2)^{\xi_3})$. Given ξ_3 , ξ_2 and γ , the upper straight line and the angle of inclination of the lower straight line to the upper straight line are fixed. As ξ_1 increases, the lower straight line slides up along the upper straight line and both coordinates of the intersection point increases.

The departure of $P_4 \setminus P_2$ from P_2 can be observed from (4.1.3) to (4.1.5) using the ξ parametrization. From (4.1.3) and (4.1.4), the approximating upper straight line differs from the actual dose response function by an amount

$$\Delta_u = \ln[1 + (\xi_1/d)^{r\xi_3}].$$

Since Δ_u decreases with increasing d , the actual curve approximates to the upper straight line better and better as $\ln d$ increases. Given an arbitrarily small $\delta > 0$, the difference Δ_u would be less than δ over the range $d_u < \ln d < \infty$.

$$\Delta_u < \delta \iff \frac{\xi_1}{(e^\delta - 1)^{1/r\xi_3}} < d$$

$$\iff \ln \xi_1 + \frac{1}{r\xi_3} \ln \frac{1}{(e^\delta - 1)} < \ln d$$

Since δ is very small, $\ln(e^\delta - 1)^{-1}$ is positive, and

$$d_u = \ln \xi_1 + \frac{1}{r\xi_3} \ln \frac{1}{(e^\delta - 1)}$$

Similarly from (4.1.3), (4.1.5), we see that the approximating lower straight line differs from the actual dose response function by an amount

$$\Delta_L = \ln\left(1 + \frac{\xi_1}{d}\right)^{r\xi_3} - \ln\left(\frac{\xi_1}{d}\right)^{r\xi_3} = \ln\left(1 + \frac{d}{\xi_1}\right)^{r\xi_3}$$

Δ_L decreases with decreasing d , hence the actual curve approximates to the lower straight line better and better as $\ln d$ decreases.

Given $\delta > 0$, the difference Δ_L would be less than δ over the range $\ln d < d_L$.

$$\Delta_L < \delta \iff \ln \left(1 + \frac{d}{\zeta_1} \right)^{r\zeta_3} < \delta$$

$$\iff \ln d < \ln \zeta_1 - (r\zeta_3)^{-1} \ln(e^\delta - 1)^{-1}$$

Again, δ is small so that $\ln(e^\delta - 1)^{-1} > 0$ and

$$d_L = \ln \zeta_1 - (r\zeta_3)^{-1} \ln(e^\delta - 1)^{-1}.$$

Note that the interval $(d_L, d_u) = (\ln \zeta_1 - (r\zeta_3)^{-1} \ln(e^\delta - 1)^{-1}, \ln \zeta_1 + (r\zeta_3)^{-1} \ln(e^\delta - 1)^{-1})$ is centered at the point (on the $\ln d$ axis) where the two approximating straight lines intersect. To the left of the interval, the actual curve approximates well to the lower straight line, the minimum difference is δ , to the right of the interval, the actual curve approximates well to the upper straight line and the maximum difference is δ . In the interval (d_L, d_u) the actual function is necessarily curved.

With this understanding on our model, we will now discuss the detectability of departure from P_2 given our experimental constraint and our intention to draw low dose inference. To recapitulate, our experimental constraint is the limit in sample size which confines us to experiment in certain observable regions where the response rate is between two numbers p_1, p_2 , $0 < p_1 < p_2 < 1$, say $p_1 = 10^{-3}$, $p_2 = 1 - 10^{-3}$. This means that we can only experiment at a dose $d \in [b_1, b_2]$ and if we are interested in drawing inference at a lower dose or response level, we will have to extrapolate. Given $p^0(d) \in P_4 \setminus P_2$, below are three

diagrams which illustrate how the experimental constraint affect the detectability of a different low dose behavior and the extrapolated result.

The three curves represent three different underlying $p^0(d) \in P_4 \setminus P_2$ with the same values of ζ_2^0 , ζ_3^0 and r^0 . The value of ζ_1^0 decreases from Diagram 7 to Diagram 9. The interval between the two crosses represent the interval (d_L, d_u) outside which the curve in the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plot approximates to a straight line to within a given δ , the interval between the two dots is the observable dose range, say $(\ln b_1, \ln b_2)$. The location of the interval (d_L, d_u) with respect to the observable region is of critical importance. Diagram 7 illustrates a case where (d_L, d_u) is completely to the right of the observable region so that the true dose response function in the observable region will closely approximate to a straight line and the data will suggest that P_2 is a right model even though there is really a difference in the high and low dose behavior. Diagram 8 illustrates a case where (d_L, d_u) overlaps with the observable dose range so that there will be a fair chance that the departure may be detected. Diagram 9 shows the case where (d_L, d_u) is completely to the left of the observable dose range so that the observable portion of the dose response function again approximates to a straight line.

These three cases indicate that extrapolation may be a very hazardous undertaking. In Diagrams 7 and 9, whereas the data may suggest very strongly what model we should use and how we should extrapolate (such as suggesting that we should use P_2 with a certain slope in the $\ln \ln(1-p(d))^{-1}$ versus $\ln d$ plot for extrapolation), the

accuracy of the extrapolated result is difficult to predict. In Diagram 7, the extrapolated result may be expected to come close to the true result; in Diagram 9, the extrapolated result may or may not come close to the true result depending on how small ζ_1^0 is and how the intersection point of the two approximating straight lines stand in relation to the dose or response rate where the inference is to be drawn.

Hence even though a large experiment may be conducted in the observable region, so long as we will have to extrapolate, there is always the danger that the change in character of the dose response function takes place below the observable region (hence remains undetected) and the apparent behavior of the dose response function in the observable region actually misleads the extrapolation.

REFERENCES

- [1] Brown, C.C. (1976), Statistical aspects of extrapolation of dichotomous dose response data. Submitted to Cancer Research.
- [2] Chambers, E.A. and Cox, D.R. (1967). Discrimination between alternative binary response models. Biometrika 54, Nos. 3 and 4, 573-578.
- [3] Chand, N. and Hoel, D.G. (1974). A comparison of models for determining safe levels of environmental agents. Reliability and Biometry, 681-700.
- [4] Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Stat. 24, 586-602.
- [5] Cramer, H. (1974). Mathematical methods of statistics. Princeton University Press, Princeton.
- [6] Crump, K.S., Hoel, D.G., Langley, C.H. and Peto, R. (1976). Fundamental carcinogenic process and their implication for low dose risk assessment. To appear in Cancer Research.
- [7] Guess, H.A. and Crump, K.S. (1976). Best-estimate low dose extrapolation of carcinogenicity data. NIEHS Extrapolation Conference Abstract.
- [8] Guess, H.A. and Crump, K.S. (1977). Can we use animal data to estimate "safe" doses for chemical carcinogens. Environmental Health Quantitative Methods, SIAM, Philadelphia.
- [9] Guess, H.A. and Crump, K.S. (1976). Low-dose-rate extrapolation of data from animal carcinogenicity experiments--analysis of a new statistical technique. To appear in Math. Biosciences.
- [10] Guess, H.A. and Crump, K.S., Maximum likelihood estimation of dose-response functions subject to absolutely monotonic constraints.
- [11] Hartley, H.O. and Sielken, R.L., Jr. (1977). Estimation of "safe doses" in carcinogenic experiments. Biometrics 33, 1-30.
- [12] Hoel, D.G., Gaylor, D.W., Kirschstein, R.L., Saffiotti, U. and Schneiderman, M.A. (1975). Estimation of risks of irreversible, delayed toxicity. J. Toxicology and Environmental Health, 1, 133-151.
- [13] Peto, R., The carcinogenic effect of chronic exposure to very low levels of toxic substances.
- [14] Whittemore, A. and Keller, J.B. (1978). Quantitative theories of carcinogenesis. SIAM Rev. 20, No. 1.