INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
UniversityMicrofilmsInternational
A Bell & Howell Information Company
300 NorthZeebRoad,AnnArbor,MI48106-1346USA
313/761-4700 800/521-0600
Multi-stage designs in dose-response studies

Spears, Floyd Martin, Ph.D.
Rice University, 1993
RICE UNIVERSITY

Multi-Stage Designs in Dose-Response Studies

by

Floyd M. Spears

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Barry W. Brown, Thesis Advisor
Adjunct Professor of Statistics

James R. Thompson, Chairman
Professor of Statistics

Katherine B. Enos
Assistant Professor of Statistics

John W. Clark
Professor of Electrical and Computer Engineering

Houston, Texas
July, 1992
Multi–Stage Designs in Dose–Response Studies

Floyd M. Spears

Abstract

Designs are explored that minimize the asymptotic variance of a single parameter in a dose-response study designed to estimate this parameter. An example is a design to find the dose producing 50% response. Uncertainty of parameter values of the dose-response curve is represented as a normal prior distribution. Because the integration of the criterion over the prior distribution is analytically untractable, numeric methods are used to find good designs.

The extension to multi–stage experiments is straightforward. The normal prior distribution coupled with the asymptotically normal likelihood yields a normal posterior distribution that is used to optimize the succeeding stage.

Simulation results suggest that the asymptotic methods are a good reflection of small sample properties of the designs, even with modest–sized experiments. If initial uncertainty of the parameters is large, two–stage designs can produce accuracy that would require a sample size fifty percent greater with a single–stage design.
Acknowledgments

I would like to thank Dr. James R. Thompson, Dr. Barry W. Brown, Dr. Katherine B. Ensor and Dr. Jack W. Clark for serving on my thesis committee. A special thanks is due Dr. Brown for his support over the last four years, as well as Dr. Stuart Zimmerman and my colleagues in the Department of Biomathematics at M.D. Anderson Cancer Center.

I would like to thank my wife, Glenda Spears, for her support and dedication throughout my years in graduate school. I also thank my parents, Shannon and Wanda Spears, and the rest of my family for the support they provided throughout my years in school.

I would like to gratefully acknowledge the support by the National Cancer Institute under grants CA16672 and CA06294 and by the Army Research Office under grant DAAL03-G-0210, as well as the computer support provided through a cooperative study agreement with IBM.
Contents

Abstract ii
Acknowledgments iii
List of Illustrations vi
List of Tables viii

1 Introduction
1.1 Designs for a Dose–Response Study 1
1.2 Hierarchy of Designs 2

2 Background
2.1 Dose–Response Models 5
2.2 Criteria 8
2.3 Single–Sample Designs 9
   2.3.1 Logistic Model with Linear Response Function 10
2.4 Two–Sample Designs 12
2.5 Designs that Consider Uncertainty of Parameter Values 13
2.6 Multi–Stage Designs 15

3 Methods
3.1 Calculation of the Criterion Value 19
   3.1.1 Asymptotic Standard Deviation 19
   3.1.2 Power of a Test 21
   3.1.3 Examples 22
   3.1.4 Calculating the Expected Criterion 23
3.2 Single-Stage Optimization Procedure ........................................ 25
3.3 Multi-Stage Optimization Procedure ........................................ 29
3.4 Multi-Stage Simulation Procedure ........................................... 30

4 Single-Stage Design ................................................................. 32
  4.1 Advantages of a Planned Design ............................................ 32
  4.2 Designs for Known Parameters ............................................. 35
  4.3 Effect of Uncertainty of Parameter Values .............................. 39
  4.4 Designs for Maximizing Power ............................................. 49

5 Multi-Stage Design ................................................................. 51
  5.1 Stage Sizes in Multi-Stage Designs ....................................... 51
    5.1.1 Optimal Stage Sizes for Two-Stage Designs ....................... 51
  5.2 Improvement over Single-Stage Design ................................ 57
  5.3 Effects of Small Sample Sizes ........................................... 61

A Expected Value of Criterion Function ...................................... 69

B Posterior Distribution of the Next Stage .................................. 72

Bibliography .............................................................................. 74

References .............................................................................. 75
Illustrations

2.1 Logistic, Probit and Complementary Log–Log Models .......... 7

4.1 Effect of Deviations from the Parameters on Precision when
   Estimating the $ED_{50}$ ........................................... 38
4.2 Effect of Uncertainty of Parameters on Designs for Estimating the $ED_{50}$ 42
4.3 Effect of Uncertainty of Parameters on Designs for Estimating the $ED_{05}$ 44

5.1 Effect of Increasing Uncertainty of $a$ on Two–Stage Designs for
   Estimating the $ED_{50}$ ........................................... 52
5.2 Effect of Increasing Uncertainty of $a$ on Two–Stage Designs for
   Estimating the $ED_{05}$ ........................................... 53
5.3 Effect of Increasing Uncertainty of $b$ on Two–Stage Designs for
   Estimating the $ED_{50}$ ........................................... 55
5.4 Effect of Increasing Uncertainty of $b$ on Two–Stage Designs for
   Estimating the $ED_{05}$ ........................................... 56
5.5 Ratio of the Asymptotic Variance of the Single–Stage Design to the
   Two–Stage Design for Estimating the $ED_{50}$ ...................... 60
5.6 Ratio of the Asymptotic Variance of the Single–Stage Design to the
   Two–Stage Design for Estimating the $ED_{05}$ ...................... 62
5.7 Asymptotic Density of $(\hat{E}D_{50} - ED_{50})$ from Two–Stage Design for
   Estimating the $ED_{50}$ ........................................... 65
5.8 Asymptotic Density of \((\hat{E}D_{10} - E\hat{D}_{10})\) from Two-Stage Design for Estimating the \(E\hat{D}_{10}\) .................................................. 67

5.9 Asymptotic Density of \((\hat{E}D_{05} - E\hat{D}_{05})\) from Two-Stage Design for Estimating the \(E\hat{D}_{05}\) .................................................. 68
# Tables

2.1 Optimal Designs for Estimating the $ED_{100-p}$ of Logistic Model with Linear Response Function .................. 11

4.1 Advantages of Planned Designs in the Linear Response Function Examples ............................................. 33

4.2 Advantages of Planned Designs in the Two-Sample and Quadratic Response Function Examples ..................... 34

4.3 Single-Stage Designs for Linear Response Function Examples, Parameters are Known .................................. 35

4.4 Single-Stage Designs for Two-Sample and Quadratic Response Function Examples, Parameters are Known ....... 36

4.5 Designs for Estimating the $ED_{50}$ with Unknown Parameters ........... 40

4.6 Designs for Estimating the $ED_{10}$ with Unknown Parameters .......... 43

4.7 Designs for Estimating the $ED_{05}$ with Unknown Parameters .......... 45

4.8 Designs for Estimating $b$ with Unknown Parameters ................. 46

4.9 Designs for Estimating $d_o - d'_o$ with Unknown Parameters .......... 47

4.10 Designs for Estimating $c$ with Unknown Parameters ................. 48

4.11 Designs for Testing $H_o : d_o - d'_o$ with Power as Criterion .......... 50

5.1 Advantages of a Two-Stage Design for the Linear Response Function Examples ........................................... 58
5.2 Advantages of a Two-Stage Design for the Two-Sample and Quadratic Response Function Examples 59
Chapter 1

Introduction

1.1 Designs for a Dose–Response Study

Dose–response studies are conducted in a variety of applications. We focus on those in biomedical research, particularly investigations relating to cancer. Some systematic increase in response rate with dose must be demonstrated to establish the effectiveness of a new treatment. In using a new treatment modality, a dose for the new method that produces the same response rate as that of the old is frequently sought. This is a common problem in radiotherapy when neutron, proton, or other beam types will be used in place of traditional gamma or photon radiation. Establishment of similar effects is necessary in order that the new treatment neither be ineffective because of too low a dose nor unacceptably toxic because of too large a dose. In some types of treatment, for example immunotherapy, there is an optimal dose. Too much of the agent is as ineffective as too little. Phase I tests of a new agent seek the dose producing a particular level of toxicity (frequently 30%). Toxicity greatly exceeding the specified level is unacceptable; the dose is thought to be biologically inactive against the cancer if the toxicity is much below the specified level.

Such studies have both ethical and economic implications. As much information as possible must be gleaned from the usually severely limited sample size available.

The traditional dose–response experiment employs a single–stage design – a fixed number of subjects are placed at predetermined dosage levels. This design is easy to administer, and can be efficient if the parameters of the dose–response relation are known with accuracy when constructing the design. A major disadvantage of
the single-stage design is its possible poor performance when the initial estimates of parameter values are wrong. This disadvantage can be circumvented by optimizing designs over a distribution representing the initial uncertainty of parameter values.

If early results indicate that a chosen single-stage design is poor, there may be both ethical and economic problems with continuing the study. The subjects in a multi-stage study are allocated in groups, allowing adaptation as the study proceeds; the number of groups and size of each group is predetermined. The doses and number of subjects at each is generally prespecified for the first stage. For succeeding stages, the design depends on the outcomes of the previous stages.

A limiting case of multi-stage studies is a sequential experiment. Subjects are allocated one at a time using the cumulative information from those previously allocated to determine the next dose level.

If the parameters of the dose-response relation are well known, then a single-stage experiment can be quite efficient. With increasing uncertainty in the knowledge of these values, multi-stage trials becomes increasingly effective as compared to single-stage trials. The information about the parameter values from initial results is used to tune the later stage of subject and dose allocation. Increasing the number of stages should increase the information obtained from the trial, so the sequential trial should be maximally effective. However, the gain diminishes rapidly with the number of stages while the administrative effort of the trial increases. Each stage of a multi-stage trial requires stopping patient accrual until all previous responses were determined. In many cancer research settings, ascertaining response requires months or years. An example is the case in which the response is the complete regression of a tumor.

1.2 Hierarchy of Designs

In comparing methods for finding designs, one must consider more than one criterion. For example, the sequential experiment is known to produce more efficient designs than both the single-stage and multi-stage experiments, but the amount of time
needed to conduct a sequential trial, if time to response is long, makes it a poor choice in many circumstances. So the costs associated with the trial must be weighed against its efficiency. The designs considered for a dose–response experiment can be ordered by their level of complexity. Using this hierarchy of designs, one can determine whether the gains obtained using a more complex design are worth the cost associated with increased complexity.

The uniform design is the least complex design considered – the dose levels are spread uniformly over some range. This design requires no planning, other than determining an acceptable range of doses, and is easy to administer. The base design, against which the others will be compared, is a uniform design with six doses spread evenly between the dose associated with a 5 percent response rate, $ED_{05}$, and the dose associated with a 95 percent response rate, $ED_{95}$. Müller's [10] results indicate that this design attains most of the efficiency of uniform designs with more dosage levels.

The single–stage design optimized with respect to the asymptotic standard deviation of the estimate for known parameter values is compared to the uniform design to determine whether substantial improvements in precision can occur from planning. In the single–stage design, the number and level of doses are fixed in advance. The single–stage design can lose most of its gain over the uniform design if the initial guesses at the parameter values are inaccurate. The need for more robust single–stage designs causes consideration of designs which are optimal with respect to the expected criterion, the criterion averaged over a prior distribution that describes the investigator's uncertainty of the parameter values. A design based on the expected criterion is more efficient than a uniform design and is robust against incorrect best estimates of the parameter values.

The multi–stage design allows adaptation based on the results of preceding stages. The techniques used to find a single–stage design with a prior distribution of the parameter values are extended to multi–stage designs. The multi–stage trial is consid-
erably more difficult to administer and is more time consuming than the single-stage trial. It must demonstrate significant improvement in precision to be considered a viable alternative to the single-stage design. Most of the gain in efficiency from adding stages to a design is attained in the first few stages [1]. Two-stage designs are compared to single-stage designs to determine if an additional stage provides a significant improvement in estimation.
Chapter 2

Background

2.1 Dose–Response Models

The purpose of dose–response experiments is to determine the probability of response as a function of dose. One might be tempted to model a dose–response relationship by supposing that \( \pi_i \), the probability of response of subject \( i \), is linearly related to the dose received \( d_i \), i.e. \( \pi_i = a + bd_i \). This is inconsistent with the laws of probability since \( \pi_i \) can range from \( (-\infty, \infty) \) in this model. There is at best a limited interpretation and range of validity for this model. One simple and effective way of avoiding this difficulty is to use a transformation \( h(\pi_i) = a + bd_i = U(d_i) \) which maps the unit interval onto \( (-\infty, \infty) \). There are numerous link functions, \( h(\pi) \), including three that are commonly used: the logistic, the probit and the complementary log–log functions.

The logistic model is popular, in part because of its natural interpretation as the logarithm of the odds ratio. The logistic function is \( h_1(\pi) = \log \left( \frac{\pi}{1-\pi} \right) = U(d) \), therefore the probability of response at dose \( d \) is

\[
P(d) = \frac{e^{U(d)}}{1 + e^{U(d)}} = \frac{1}{1 + e^{-U(d)}},
\]

where \( U(d) \) is the dose–response function. The probability distribution function is symmetric about the median dose, so the cumulative distribution function has its inflection point at the \( ED_{50} \).

Another model is the probit. The link function of the probit model is the inverse cumulative distribution function of the standard normal, \( h_2(\pi) = \Phi^{-1}(\pi) = U(d) \). The resulting probability of response is

\[
P(d) = \Phi\{U(d)\}.
\]
The probit model can be closely approximated by the logistic model, which was developed as an approximation to the normal cumulative distribution function. As is the case for the logistic function, its probability distribution function is symmetric about the median dose.

The complementary log-log model is based on the assumption that a response is the occurrence of a point in a Poisson process with rate \( \lambda \). The complementary log-log link function is \( h_3(\pi) = \log \left\{ -\log (1 - \pi) \right\} = U(d) \), with the associated probability of a response at dose \( d \) being

\[
P(d) = 1 - e^{-e^{U(d)}}.
\]

The natural counterpart of the complementary log-log model is the log-log model which has the link function \( h_4(\pi) = -\log \left\{ -\log (\pi) \right\} = U(d) \). The log-log function is less common in practice than the complementary log-log model. These two models are not symmetric, but are related via \( h_3(\pi) = -h_4(1 - \pi) \). The inflection point for the complementary log-log function is located at \( P(d) = 1 - e^{-1} \).

Figure 2.1 is a plot of the maximum likelihood fits of these three models to a data set from Cox and Snell’s text *Analysis of Binary Data* [4]. The data has 5 dose levels of a treatment with 30 subjects at each level. The logistic and probit models produce almost identical fits to the data. The complementary log-log model differs from the other models, most noticeably in the tails.

The dose-response function, \( U(d) \), is frequently assumed to be either linear or quadratic. The linear response function is used if there is a monotone relationship between dose and response, and can be written as

\[
U(d) = a + bd = b(d - d_o)
\]

where \( a \) is the intercept, \( b \) is the slope and \( d \) is either the dose or the log-dose. The alternative parameterization centers the doses around \( d_o \), the dose associated with 50% response.
Figure 2.1  The maximum likelihood fits of the logistic, probit and complementary log-log models with a linear response function to data from The Analysis of Binary Data [4].
The quadratic dose–response function is used if there is a dose yielding a maximum probability of response. Doses away from this one in either direction produce a decrease in the probability of response. The quadratic response function can be written as

\[ U(d) = a - b(d - c)^2 = a' + b'd + c'd^2. \]

The first parameterization is preferred since it centers the doses around the dose associated with the maximum of the quadratic and has interpretable parameters. The height and the width of the quadratic are controlled by \( a \) and \( b \) respectively, while \( c \) is the dose associated with the maximum of the quadratic.

Changes in the parameters of the linear and quadratic response functions affect the probability of response in an obvious manner. For the linear response function, increasing the slope, \( b \), results in a steeper probability response curve. Changes in the location parameter, \( a \), shift the probability response curve horizontally by the amount of the change, but in the opposite direction. For the quadratic response function, increasing \( a \) results in a larger maximum for the quadratic, but does not effect the slope. Changes to the centering parameter, \( c \), shift the probability response curve an amount equal to the change, but in the opposite direction. Increasing \( b \) causes an increase in the slope of the probability response curve.

### 2.2 Criteria

The outcome of a dose–response study can be largely affected by the design chosen. The criterion for selecting a design can be divided into two groups. One group includes criteria that perform well, regardless of the purpose of the experiment; the emphasis of such a criterion is placed on obtaining good estimates of the parameters. An example of such a criterion is the determinant of the asymptotic variance–covariance matrix. Minimization of the determinant is known as D–optimality in linear design. Another
example of such a criterion is the trace of the asymptotic variance-covariance matrix, whose minimization is known as A-optimality in linear design.

Although it is important in some cases to obtain a design which is robust with respect to the purpose of the experiment, the primary purpose of many dose-response experiments is specific, i.e. to estimate or test some quantity of interest. As one example, biomedical trials are conducted to find the $ED_{100,p}$, the dose of a drug which will result in a certain proportion $p$ of the subjects responding. As another example, trials are conducted to determine if there is a monotone relationship between dose and response; this example tests the null hypothesis that $b = 0$ for the linear response function.

The second group of criteria include those which place emphasis on maximizing the precision with which a quantity is estimated. These criteria produce designs which perform very well for specific purposes, but can be nonrobust to changes in the purpose of the experiment. An example is minimizing the asymptotic variance of the estimate of a quantile of response. The optimal design for estimating a quantile of response differs greatly for different quantiles. Another example of a criterion for a specific purpose is the power of a test. Maximizing this criterion assures a researcher of the maximum probability of rejecting a false null hypothesis. An alternative criterion to the power of a test is the asymptotic standard deviation of the quantity being tested.

2.3 Single-Sample Designs

Single-sample experiments are conducted with a single treatment being administered to a group of subjects. The purpose of many such experiments is to estimate or test some quantity; emphasis is placed on maximizing the precision of the estimate for the number of subjects available. A common criterion for such a study is the asymptotic variance of the estimate. This requires calculation of the asymptotic variance-covariance matrix of the parameter estimates.
In a dose–response experiment, the responses $y_1, y_2, \ldots, y_n$ are observed values of independent random variables $Y_1, Y_2, \ldots, Y_n$ such that $Y_i$ is binomial with sample size one and the probability of response is $P(d_i)$; $d_i$ is the dose administered to the $i^{th}$ subject. The likelihood function is the product of the binomial response probabilities

$$L = \prod_{i=1}^{n} P(d_i)^{Y_i}(1 - P(d_i))^{1-Y_i}$$

yielding the asymptotic variance–covariance matrix

$$\text{var}(\hat{\theta}) = \left[ -E_0 \frac{\partial^2}{\partial \theta^2} \ln L \right]^{-1}.$$  

The asymptotic variance of a parameter estimate, $\hat{\theta}_i$, is the associated diagonal element of the variance–covariance matrix, $\left[ \text{var}(\hat{\theta}) \right]_{[i,i]}$. The asymptotic variance of a function of the parameters estimates, $f(\hat{\theta})$, is obtained using propagation of error (see equation 3.3) and the asymptotic variance–covariance matrix because the maximum likelihood estimates are asymptotically normal [13].

2.3.1 Logistic Model with Linear Response Function

A number of researchers have addressed the problem of finding an optimal design for estimating the $ED_{100 \cdot p}$. Hoel and Jennrich [7] show using [Chernoff, 1953] that the optimal number of dose levels for estimating the $ED_{100 \cdot p}$ is at most the number of parameters in the model. This implies that there are no more than two doses in the optimal design for estimating a quantile of response for a logistic model with a linear response function.

Tosh and McLeish [15] use this result to solve explicitly for the optimal design for estimating a quantile of response using the logistic model with a linear response function. The number of doses in the optimal design is dependent on whether or not the quantile of response estimated is central or extreme. The optimal design for estimating a central quantile of response, $ED_{100 \cdot p} \in (r_1, r_2)$, is a single dose at the quantile (the dose levels, $r_1$ and $r_2$, are determined by the model). For estimating
an extreme quantile of response, \( ED_{100-p} \in (-\infty, r_1] \cup [r_2, \infty) \), a two dose design is optimal with subjects at \( r_1 \) and \( r_2 \). The number of subjects at \( r_1 \) and \( r_2 \) in this design is dependent on the particular quantile estimated. For the logistic model with a linear response function, \((r_1, r_2) = (ED_{8.3}, ED_{91.7})\). Table 2.1 displays the asymptotically optimal designs for estimating several quantiles of response, assuming this model.

<table>
<thead>
<tr>
<th>Quantile Estimated</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ED_{01} )</td>
<td>76.1/23.9</td>
<td>( ED_{8.3}/ED_{91.7} )</td>
</tr>
<tr>
<td>( ED_{05} )</td>
<td>90.7/9.3</td>
<td>( ED_{8.3}/ED_{91.7} )</td>
</tr>
<tr>
<td>( ED_{10} )</td>
<td>100</td>
<td>( ED_{10} )</td>
</tr>
<tr>
<td>( ED_{50} )</td>
<td>100</td>
<td>( ED_{50} )</td>
</tr>
</tbody>
</table>

**Table 2.1** The analytically optimal designs, found by Tosh and McLeish [15], for estimating quantiles of response, \( ED_{100-p} \), of a logistic model with a linear response function.

The work of Tosh and McLeish on the logistic model was extended to the complementary log–log model with a linear response function. The results for this model are very similar to those for the logistic model. The optimal design for estimating a central quantile has a single dose at the quantile; the optimal design for an extreme quantile has subjects placed at the doses, \( r_1 \) and \( r_2 \). For the complementary log–log model, the doses \((r_1, r_2) = (ED_{11,8}, ED_{97,1})\) are close to symmetrical about the inflection point, \( ED_{100-(1-c^{-1})} \).

Although the analytically optimal design for estimating a central quantile has all the subjects at the quantile, a design with one dose does not allow well-defined estimates for both parameters of the linear response function. Singularity problems are encountered for any design with fewer dosage levels than parameters of the model.

Müller [10] looks at the optimal number of dose levels when estimating the \( ED_{50} \) from a different perspective. He notes that a range of valid doses can usually be
specified before a study begins. The selection of the doses to be administered can then be formalized by defining a design density on the given range. Doses are chosen at equally spaced quantiles of this density. Only designs with an equal allocation of subjects are considered, therefore the design is determined by the number of dose levels chosen.

This class of designs allows for a large variety. A uniform design density has equally spaced doses across the range of valid doses. A normal design density concentrates more doses in the center, while an exponential design density places more subjects at lower dose levels. The choice of design density has little effect on the optimal number of dose levels. The maximum number of dose levels possible is the best choice, except when the range of doses is large compared to the scale parameter of the linear model, i.e. large values of the slope, $b$, compared to the range of doses for the linear response function.

Requiring the subjects to be allocated equally to the various doses is a serious limitation. An example is a three dose design for estimating the $ED_{50}$. Allocating the subjects equally places 66% of the subjects at dosage levels other than the optimum, placing all the subjects at the $ED_{50}$. This three dose design is inefficient. If subjects were not equally allocated, a majority of the subjects could be placed at the center dose, resulting in a design that is close to optimal.

2.4 Two-Sample Designs

There are two distinct forms of a two-sample experiment. One forms treatment groups by administering two different treatments to subjects with similar makeup. The purpose of this study is to show whether or not there is a difference in the two treatments. A two-sample experiment is also formed by administering a single treatment to two distinctly different subject groups. The purpose of this study is to determine if there is a difference in the way the two groups respond to the single treatment.
Two treatment groups are compared in a dose–response experiment by comparing their probability response curves. For the relatively small sample sizes used in many radiobiology studies, the slope of the two curves is thought to be close enough to assume that it is the same for both groups. In this situation, one can test for a difference between the two curves by looking at the difference between the doses producing 50% response for the two treatment groups, $d_o - d_o'$. 

One criterion for selecting a design in this situation is the asymptotic standard deviation of the estimate of the difference, $d_o - d_o'$; minimizing this criterion maximizes the precision of the test statistic. The asymptotic standard deviation of $\hat{d}_o - \hat{d}_o'$ can be calculated using the single-sample methods described in section 2.3 with the parameterization, $\theta = (b, d_o, d_o')$, where $b$ is the slope common to both groups (see example 2b in section 3.1.3).

An alternative criterion is the power of a test of the difference between treatment groups; the alternative hypothesis is $H_a : d_o - d_o' \neq 0$. Maximizing the power maximizes the probability that a researcher will reject the null hypothesis when the alternative hypothesis is true. The probability of detecting a difference between parameters is calculated using the asymptotic distribution of twice the log-likelihood, a chi-square. The chi-square is central under the null hypothesis and noncentral under the alternative hypothesis. The details of this power calculation are discussed in section 3.1.2.

### 2.5 Designs that Consider Uncertainty of Parameter Values

The single-sample and two-sample designs discussed in sections 2.3 and 2.4 assume that the parameters are known. Of course, there is no need to conduct an experiment if the parameters are known a priori, so some "best guess" at the true parameter values must be used. These designs are good as long as the parameters are estimated with a fair amount of accuracy; they can be quite inefficient for values far from the truth. This type of design was coined a "locally optimal" design by [Chernoff, 1953].
Designs based on specific parameter values can lose significant precision in estimation if the parameter values differ, even slightly, from the truth. This can be compensated for by larger sample sizes.

One alternative to the "locally optimal" design is to use a prior distribution to describe the investigator's uncertainty of the parameter values. The criterion can then be averaged over the prior distribution to obtain its expected value. Optimizing with respect to the expected criterion provides a design that is more robust against incorrect best estimates of the parameter values.

Chaloner [3] explores designs optimal with respect to the expected criterion. She considers designs which minimize the asymptotic expected posterior variance of the quantities of interest. A weighted trace of the inverse of the expected information matrix is averaged over the prior distribution and the resulting value is minimized for the optimal design. Equally weighting each member of the diagonal of the information matrix results in a general criterion. Placing a large weight on the diagonal of the information matrix associated with a parameter of interest creates a specific criterion.

The other criterion examined by Chaloner is the average over the prior distribution of the log of the determinant of the information matrix, which is maximized in a design. This corresponds, approximately, to the maximum expected increase in Shannon information provided by the experiment. This is a general criterion, placing emphasis on estimating all of the model parameters. Chaloner chooses a uniform prior distribution on parameters $b$ and $d_o$.

Chaloner finds designs optimal with respect to the expected criterion for the $ED_{50}$ and the $ED_{95}$, so both central and extreme quantiles are considered. A third example considered is estimating some dose $x$ so that $P(x) = p$ where $p$ has a uniform distribution over $[-1, 1]$. This range of $\text{logit}(P(x))$ corresponds to $p$ ranging from 0.27 to 0.73. In all three of these cases, the optimal Bayesian design is similar to the locally optimal design when the variance of the prior distribution is relatively small. As the
prior variance increases, the design based on the expected criterion has more dose levels.

According to Chaloner, the best design for the \( ED_{50} \) with a small prior variance is two doses centered about the prior mean of \( d_0 \). More doses are added in two's centered around the prior mean of \( d_0 \) as the variance of the prior distribution increases. When estimating the \( ED_{95} \), more dose levels are added to the design as the prior variance increases, with most of these doses located in the tail right tail.

Recall from section 2.3 that Müller [10] looks at a class of designs with doses placed at equally spaced quantiles of a design density and an equal number of subjects at each dose. This class of designs has as its optimum one with the maximum number of dose levels possible. Chaloner does not restrict the distribution of subjects and finds that the optimal designs do not have an equal distribution of the subjects among the doses. She finds that the designs have more doses as the uncertainty of the parameter values increases. Efficiency calculations performed by Chaloner on 9 different combinations of uniform priors for \( b \) and \( d_0 \), using minimum variance as the criterion, reveal that for the worst combination, 480 times as many observations were needed for the "locally optimal" design to match the design that takes into account uncertainty in the parameter values.

2.6 Multi–Stage Designs

A multi–stage design is a practical compromise between a single–stage design and a sequential design. Unlike the sequential design which requires analysis and design adjustments after each observation, subjects in a multi–stage design are treated in groups (stages). The number of stages and the allocation of subjects to each stage is decided in advance. The design chosen at each stage is dependent upon the results from the preceding stages.

Both the time to administer a study and the complexity of a study increase with the number of stages. Most of the improvement in terms of efficiency of a multi–stage
design over a single-stage design is in the first few stages added. A two-stage design can greatly improve efficiency with a relatively small increase in time.

Abdelbasit and Plackett [1] compare single-stage designs to multi-stage designs that have the same number of subjects. They examine one and two parameter logistic models with up to five stages in the design. They choose the the determinant of the information matrix as their criterion; maximizing the determinant corresponds to D-optimality.

Multi-stage designs are almost always more efficient than single-stage designs; the exception being if the initial estimate of the parameters is very close to truth. When the initial estimate is extremely accurate, the efficiency is slightly less than 1.0. The efficiency increases as the sample size increases and as the initial estimate moves further from the truth. For large sample sizes, the efficiency is at least 1.0 everywhere. So unless the parameter values are known quite accurately, a multi-stage design is guaranteed of performing at least as well as, and generally much better than, a single-stage design.

Tosh and McLeish look at sequential designs for minimizing the asymptotic standard deviation when estimating a quantile of response using the logistic model with a linear response function $U = b(d - r)$, where $r$ is the quantile of response being estimated [15]. If $P(d)$ is the probability of a response at a given dose, the expected information matrix for the first $n$ observations is

$$I_n(r, b) = \sum_{i=1}^{n} w(d_i) \begin{bmatrix} b^2 & -b(d_i - r) \\ -b(d_i - r) & (d_i - r)^2 \end{bmatrix},$$

where $d_i$ represents the dose levels administered and

$$w(d_i) = \frac{(P'(d_i))^2}{P(d_i)(1 - P(d_i))}.$$

The next dose level is selected in such a way as to minimize the variance of the maximum likelihood estimator of $r$. After the first $n$ observations, $\hat{r}_n$ and $\hat{b}_n$ are the maximum likelihood estimates of $r$ and $b$. The variance of the maximum likelihood estimate of $r$ at the next dosage level $d_{n+1}$ is the $(1,1)$ entry of the inverse of
\[ I_n(\hat{r}_n, \hat{b}_n) + w(d_{n+1}) \begin{bmatrix} \hat{b}_n^2 & -\hat{b}_n(d_{n+1} - \hat{r}_n) \\ -\hat{b}_n(d_{n+1} - \hat{r}_n) & (d_{n+1} - \hat{r}_n)^2 \end{bmatrix}, \]

the expected information matrix. This sequential method is easily adapted to a multi-stage design. The information matrix for the next stage is found by adding the product of the information for the next dose level and the number of subjects to receive that dose. It is written

\[ I_k(\hat{r}_k, \hat{b}_k) + n_k w(d_{k+1}) \begin{bmatrix} \hat{b}_k^2 & -\hat{b}_k(d_{k+1} - \hat{r}_k) \\ -\hat{b}_k(d_{k+1} - \hat{r}_k) & (d_{k+1} - \hat{r}_k)^2 \end{bmatrix}, \]

where \( I_k(\hat{r}_k, \hat{b}_k) \) is the expected information matrix, \( \hat{r}_k \) and \( \hat{b}_k \) are the maximum likelihood estimates, and \( n_k \) is the number of subjects for stage \( k \).

Tosh and McLeish show that the design found using their sequential procedure approaches the optimal design for the logistic model in which the parameter values are known. From section 2.3.1, the locally optimal design for the logistic model estimating \( r \) is either a one point design when \( r \) is a central value or a two point design when \( r \) is an extreme value. One of the major advantages of this sequential procedure is that there is no need to know in advance how many dose levels are in the optimal design. All that is needed is the ability to compute the expected information matrix at each step. As more information is collected, the sequential design approaches the optimal design.

Their minimum variance procedure was compared to the logit-MLE model since it is shown to achieve asymptotic efficiency by Wu [16]. The logit-MLE uses the first \( n \) observations to compute the maximum likelihood estimates for \( r \) and \( b \). An estimated probability response curve is constructed using the current maximum likelihood estimates and the next design point is chosen to be \( \hat{r}_n \) from this curve.

The minimum variance method is shown to be superior to the logit-MLE method when estimating extreme quantiles and very competitive when estimating central quantiles. The design points obtained using the minimum variance method surround
the true value. The design points obtained using the logit–MLE method approach the true value from the direction in which the original maximum likelihood estimates started; this can result in a biased estimate.
Chapter 3

Methods

A number of different models are used for dose–response data (see section 2.1). The design optimization techniques described in this chapter can be applied to any model. The logistic model is studied exclusively; it is a popular model, expressible in terms of elementary functions which simplifies construction of the asymptotic variance–covariance matrix. All but one of the examples are for the linear response function; the exception is for the quadratic response function.

The criteria for selecting a design for a dose–response study divide into two groups (see section 2.2). One group of criterion produces designs which perform well, regardless of the purpose of the study. These designs can be improved in particular cases by choosing a criterion representing the purpose of the experiment.

The purpose of many dose–response studies is to estimate or test a specific quantity. The optimal design for such a study maximizes the precision with which the quantity of interest is estimated. Two such criteria studied exclusively in the succeeding sections are the asymptotic standard deviation of an estimate and the power of a test.

3.1 Calculation of the Criterion Value

3.1.1 Asymptotic Standard Deviation

The asymptotic standard deviation of a parameter or a function of the parameters is derived from the asymptotic variance–covariance matrix. As discussed in section 2.3, the responses $y_1, y_2, \ldots, y_n$ in a dose–response experiment are assumed to be the
observed values of independent random variables, binomially distributed with probability of response $P(d_i)$, where $d_1, d_2, \ldots, d_n$ are the dose levels administered. The likelihood function is the product of the independent binomial response probabilities,

$$L = \prod_{i=1}^{n} P(d_i)^{y_i}(1 - P(d_i))^{1-y_i},$$

and the score function is

$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^{n} \frac{y_i - P(d_i)}{P(d_i)(1 - P(d_i))} \frac{\partial U(d_i)}{\partial \theta}.$$ 

Evaluating $\left[ \frac{\partial^2 \ln L}{\partial \theta^2} \right]$ at the expected response, $E(Y_i) = P(d_i)$, leads to the expected information matrix

$$I(\theta) = \left[ -\sum_{i=1}^{n} w_i(d_i) \frac{\partial U(d_i)}{\partial \theta} \right],$$

where $w_i = P(U_i)(1 - P(U_i))$ for the logistic model. The expected information matrix for the logistic model with a linear response function is

$$I_l(a, b) = \sum_{i=1}^{n} P(U_i)(1 - P(U_i)) \begin{bmatrix} 1 & d_i \\ d_i & d_i^2 \end{bmatrix}. \quad (3.1)$$

where $d_i$ is the dose administered to the $i^{th}$ observation. For the logistic model with a quadratic response function, the expected information matrix is

$$I_q(a, b, c) = \sum_{i=1}^{n} P(U_i)(1 - P(U_i)) \begin{bmatrix} 1 & -(d_i - c)^2 & 2b(d_i - c) \\ -(d_i - c)^2 & (d_i - c)^4 & -2b(d_i - c)^3 \\ 2b(d_i - c) & -2b(d_i - c)^3 & 2b(d_i - c)^2 \end{bmatrix}. \quad (3.2)$$

The expected information matrix is inverted to obtain the asymptotic variance–covariance matrix. The asymptotic variance of a parameter is the associated diagonal element of the asymptotic variance–covariance matrix. The asymptotic variance of a function of the parameters, $f(\theta)$, is calculated using propagation of error [13].
\[
\text{var}\{f(\hat{\theta})\} = \sum_{i=1}^{m} \left( \frac{\partial f}{\partial \theta_i} \right)^2 \text{var}(\hat{\theta}_i) + 2 \sum_{1 \leq i < j \leq m} \left( \frac{\partial f}{\partial \theta_i} \right) \left( \frac{\partial f}{\partial \theta_j} \right) \text{cov}(\hat{\theta}_i, \hat{\theta}_j)
\]

where \(\text{var}(\hat{\theta}_i)\) and \(\text{cov}(\hat{\theta}_i, \hat{\theta}_j)\) are elements of the asymptotic variance–covariance matrix.

### 3.1.2 Power of a Test

The power of the test of a parameter value is another criterion for selecting a design. Maximizing the power assures the maximum probability of rejecting a false null hypothesis. Designs with maximum power are compared to designs which minimize the asymptotic standard deviation of the estimate of the quantity tested.

The power of a test can be calculated by decomposing the parameters, \(\theta\), into \((\psi, \lambda)\) where \(\psi\) includes the parameters to be tested and \(\lambda\) is composed of nuisance parameters, whose values are not of interest. Twice the log of the likelihood ratio has a limiting noncentral chi-square distribution with noncentrality parameter \(W\) and degrees of freedom equal to the number of parameters being tested. The noncentrality parameter, \(W\), is shown in Cox and Hinkley's book, *Theoretical Statistics* [5], to be

\[
W = (\hat{\psi} - \psi_0)\{i_{\psi\psi}(\psi_0, \lambda) - i_{\psi\lambda}(\psi_0, \lambda)i_{\lambda\lambda}^{-1}(\psi_0, \lambda)i_{\psi\lambda}^T(\psi_0, \lambda)\}(\hat{\psi} - \psi_0) + o_p(1),
\]

where the information matrix \(I(\psi_0, \lambda)\) has been partitioned according to the partitioning of \(\theta\) and takes the form

\[
I(\psi, \lambda) = \begin{bmatrix}
i_{\psi\psi}(\psi, \lambda) & i_{\psi\lambda}(\psi, \lambda) \\
i_{\lambda\psi}(\psi, \lambda) & i_{\lambda\lambda}(\psi, \lambda)
\end{bmatrix}.
\]

The chi-square is central under the null hypothesis, so the critical value, \(c_\alpha\), associated with the significance level, \(\alpha\), is obtained via a cumulative chi-square calculation. The power of a test against the null hypothesis is the probability in the critical region under the limiting distribution of the alternative hypothesis, a noncentral chi-square distribution with non-centrality parameter \(W\). The critical region is defined by the critical value \(c_\alpha\) from the null hypothesis.
3.1.3 Examples

Enumerated below are several examples of quantities estimated or tested in cancer investigations. The results in chapters 4 and 5 are from single-stage and multi-stage designs for estimating these quantities. The asymptotic standard deviation of the estimate of the quantity of interest is considered as a criterion for all of the examples. The power of the test described in example 2b is also considered as a criterion.

1. **Quantile of response,** \( E_{D100,p} \) – the dose corresponding to probability of response, \( p \). Designs for estimating quantiles of response for a logistic model with a linear response function divide naturally into extreme and central quantile designs (section 2.3). The specific examples considered include a central quantile, \( E_{D_{50}} \), and an extreme quantile, \( E_{D_{05}} \), as well as a central quantile close to being extreme, \( E_{D_{10}} \). The quantile of response for this model is a function of the parameters

\[
q(a, b) = E_{D100,p} = \frac{\ln \left( \frac{p}{1-p} \right) - a}{b}.
\]

The asymptotic variance of the maximum likelihood estimate of \( q(a, b) \) is obtained using propagation of error (equation 3.3)

\[
\text{var}\{q(\hat{a}, \hat{b})\} = \frac{\text{var}(\hat{a})}{\hat{b}^2} + \frac{\text{var}(\hat{b})(\ln \left( \frac{p}{1-p} \right) - \hat{a})^2}{\hat{b}^4} + \frac{2\text{cov}(\hat{a}, \hat{b})(\ln \left( \frac{p}{1-p} \right) - \hat{a})}{\hat{b}^3}
\]

2. **Model parameters.** Designs are explored for estimating or testing the model parameters for both one-sample and two-sample trials.

   (a) **One-Sample** Testing the null hypothesis, \( H_o : b = 0 \), for the linear response function is a method for testing whether the response is monotonically dependent on dose. The asymptotic variance of \( \hat{b} \) is derived from equation 3.1

\[
\text{var}(\hat{b}) = \left[ I^{-1}(\hat{a}, \hat{b}) \right]_{2,2}.
\]
(b) **Two-Sample** Testing the null hypothesis, \( H_0 : d_o - d'_o = 0 \), for the linear response function is equivalent to testing for a difference between dose–response relationships in the two groups, assuming a common slope, \( b \). The asymptotic variance of \( \hat{d}_o - \hat{d}'_o \) is calculated using equations 3.1 and 3.3,

\[
\text{var}(\hat{d}_o - \hat{d}'_o) = \text{var}(\hat{d}_o) + \text{var}(\hat{d}'_o) - 2\text{cov}(\hat{d}_o, \hat{d}'_o).
\]

The power of the test of \( H_0 \) is calculated using the methods described in section 3.1.2.

3. **Optimal dose for a quadratic response function** – the dose resulting in the maximum probability of response for the quadratic model. In a quadratic dose–response relationship, doses varying in either direction from the optimum are equally less effective. The quadratic response function is parameterized as \( U(d) = a - b(d - c)^2 \). This parameterization yields a ready interpretation for the parameters, as discussed in section 2.1. The asymptotic variance of \( \hat{c} \) is derived from equation 3.2

\[
\text{var}(\hat{c}) = \left[ I^{-1} \left( \hat{a}, \hat{b}, \hat{c} \right) \right]_{[3,3]}.
\]

3.1.4 **Calculating the Expected Criterion**

The criterion described in the previous section, the asymptotic standard deviation of an estimate and the power of a test, are for a design at specific parameter values; calculation of these criterion is straightforward. Designs based on a criterion calculated at a specific parameter value can be inefficient at different parameter values. Designs which are more robust to incorrect estimates of the parameter values are obtained using the expected criterion calculated by averaging the criterion over a prior distribution which describes the uncertainty of the parameter values.
The expected criterion is analytically untractable, so must be approximated numerically. For dose–response functions with more than one parameter, the prior distribution is multivariate and can have correlations between the parameters. A spectral decomposition of the variance–covariance matrix allows the parameters to be transformed into an uncorrelated space; this is known as an orthogonalization. Algebraically manipulations of the expected criterion in the uncorrelated parameter space put it into a form approximatable using Hermitian formulae. This approximation is described in detail in Appendix A.

Approximating the expected criterion with the Hermitian formula involves evaluating the criterion at various parameter values. Parameter values in the tail of the prior distribution can produce an extremely large asymptotic standard deviation of the estimate. If a set of such parameter values is used in the numerical approximation procedure, smaller asymptotic standard deviations of the estimate from parameters located in the center of the prior are overwhelmed. This problem is not just theoretical, as distributions of parameter values from real data sets have produced estimates with essentially infinite asymptotic standard deviation for parameter values less than two standard deviations from the mean.

Using parameter values which produce large asymptotic standard deviation of the estimate for approximating the expected criterion causes too much emphasis in the optimization procedure to be placed on avoiding bad designs. It may be preferable to find good designs for more probable parameter values.

An alternative criterion to the asymptotic standard deviation is the reciprocal asymptotic standard deviation of the estimate. Using this criterion, parameter values causing essentially infinite asymptotic standard deviation of the estimate have a benefit of 0. Averaging the reciprocal of the asymptotic standard deviation over the prior distribution produces the expected harmonic mean of the asymptotic standard deviation. Optimizing with respect to the minimum reciprocal of the expected harmonic mean of the asymptotic standard deviation places emphasis on obtaining
optimal designs under favorable conditions. To avoid the problems described above, the reciprocal of the expected harmonic mean of the asymptotic standard deviation of the estimate,

\[ sd^*(\theta) = \frac{1}{\int(sd(\theta))^{-1}d\theta} \]

will be used instead of the expected asymptotic standard deviation of the estimate.

Routines have been written in FORTRAN and S-PLUS* which calculate the asymptotic variance-covariance matrix for parameters of the logistic model. This matrix is used to calculated the asymptotic standard deviation of parameters or functions of the parameters or the power of a test of the difference between parameters. This includes all of the examples from section 3.1.3.

Given a design, criterion can either be evaluated at specific parameter values or averaged over the prior distribution on the parameter values to obtain the expected criterion. The procedures used to find single-stage and multi-stage designs are described in the following sections.

### 3.2 Single-Stage Optimization Procedure

The problem of finding a single-stage design for a dose-response study is approached as a sequence of minimization problems instead of as a single minimization problem. The number of dose levels in the design is unknown, so designs with a different number of dose levels are compared. One approach to finding a single-stage design is to start with each subject at a different dose level and optimize the doses. Doses are removed if they are determined to be unnecessary by the optimization procedure. This approach is a high dimensional optimization problem and might be infeasible.

An alternative approach to finding a single-stage design is to start by finding a design with the minimum number of doses necessary to fit the model. If additional

---

*Analysis were carried out in S-Plus Version 3.0 (Statistical Sciences, Inc., 1991)*
doses result in a more efficient design, the doses are added and an improved design is obtained. The first design is the simplest possible design; more complex designs are selected only if they provide significant improvement over simple designs.

The following algorithm is for finding a single-stage design for a specific model and criterion. If there is some uncertainty of the parameter values, the same optimization procedure is performed using the expected criterion. This algorithm is also used to find the design for each stage of a multi-stage design.

1. **Scan for initial design.** The initial design is found by scanning the criterion over a grid of equally spaced doses with the subjects allocated equally between them. The designs in the scan procedure are limited to those having the same number of doses as parameters in the model. A two dimensional scan is conducted for the quadratic response function by setting one of the doses to the optimal dose level, \( c \), and scanning over the other two doses. The three dimensional scan is avoided since the number of criterion evaluations goes up exponentially with the number of dimensions.

The initial design contains the combination of doses from the scan procedure yielding the best criterion value.

2. **Optimize designs for a specific number of doses.** A series of optimization procedures are performed on the initial design. David Gay's minimization routine, RMNFB\(^\dagger\), is the optimization routine used in each procedure. This sequence of optimization procedures is performed on designs with a fixed number of dose levels. The first procedure listed below only applies to the two-sample model; the remaining procedures apply to all of the models.

   (a) (Only performed with two-sample data) The initial design has identical doses in the two groups with the subjects allocated equally to the doses.

\(^\dagger\)Upgraded version of David Gay's unconstrained minimization routine described in *ACM The Transactions in Mathematical Software*. 
The criterion is optimized with respect to dose levels with the constraints that the two groups have the same dose levels and the subjects are equally allocated.

(b) The criterion is optimized with respect to the dose levels while the subjects are constrained to be equally allocated. For future reference, the final design from this procedure is design $A$.

(c) The dose levels from procedure 2b are held constant while the criterion is optimized with respect to the number of subjects at each dose.

(d) Unconstrained optimization allows the doses and the subject allocation to vary as the criterion is optimized.

This sequence of optimization procedures obtains a design, $D$, for the current number of doses. If any two doses in $D$ are nearly identical, they are combined into a single dose, and the new design is accepted as optimal. If the subject allocation at a particular dose is less than $10^{-5}$, that dose is dropped from the design and the new design is accepted as optimal.

3. Check for improvement with additional doses. To determine if an additional dose level results in a better design, one dose level is added to the design found in optimization procedure 2b, $A$; call this new design $A_1$. If each dose in $A$ has $m$ subjects, the new dose receives $\frac{m}{2}$ subjects, equally taken form each dose level in design $A$.

(a) The criterion is evaluated for design $A_1$ as the new dose is scanned over the grid of equally spaced doses from procedure 1. The dose from the scan associated with the best design, $A_1$, is used as the starting value for a function minimization routine which refines the value. The value from this routine is the new dose for the initial design, $A_1$. 
(b) If there is no additional dose level so that $A_1$ is better than $D$, the number of subjects at the new dose level is cut in half; the other half is allocated equally to the original doses. The one-dimensional scan procedure from 3a is repeated. Halving continues until an improvement is found or until the number of subjects at the new dose level is reduced to less than one percent of the total. In the later case, the optimization procedure is stopped and $D$ is accepted as the best design.

(c) If a design with an additional dose is found to be better than $D$, step 2 is repeated using the new design, $A_1$ as the starting point.

An important step in the sequence of optimization procedures is determining if additional dose levels result in a more efficient design. A more complex design, one with more dose levels, is only desirable if it is significantly more efficient than those with fewer doses.

Chaloner shows for similar criteria that if placing some $\epsilon$ of the subjects of a design, $D_0$, at an additional dose level does not result in an improved design for any dose, then design $D_0$ cannot be improved on by adding dose levels [3]. This method is altered slightly in step 3 since comparing the final design, $D$, to a design with an additional dose level often results in several dose levels being added to $D$ with little improvement in the criterion. The alteration is to check for the need for additional doses using the design with equal subject allocation, $A$. This reduces the number of dose levels added which don’t significantly improve on design $D$.

An alternative approach to checking for additional doses is to repeat the optimization procedures for each additional dose level and choose the best design. The cost, in terms of computer intensity, of this alternative is unacceptable.
3.3 Multi-Stage Optimization Procedure

Finding an optimal multi-stage design is approached as a sequence of single-stage optimization procedures; the prior distribution is updated after each stage to include the additional information. The number of stages and the allocation of subjects to each stage are fixed. The design of each stage is found using the single-stage optimization procedure described in the previous section. The design for the first stage is fixed by the prior distribution since there is no further information. Designs for each of the remaining stages are found by averaging the criterion over the posterior distribution from the preceding stage. For comparison with the single-stage design, the criterion for a multi-stage design can be calculated for specific parameter values using all the dose levels and subjects from the individual stages.

The prior distribution of the parameters is \( \pi(\theta) \sim N(\theta_\pi, \Sigma_\pi) \). The expected criterion of a multi-stage design is the criterion averaged over \( \pi(\theta) \). The expected criterion is analytically untractable, but it can be approximated using numerical integration. This approximation is accomplished using the Hermitian formulae, which requires evaluating the criterion at numerous integration points.

To evaluate the criterion at an integration point, \( \theta_o \), the multi-stage design for \( \theta_o \) is constructed. The design for the first stage is fixed by \( \pi(\theta) \), so it is only calculated once. Stages after the first differ for different integration points; the criterion is averaged over the posterior distribution from the previous stage instead of the prior distribution.

To calculate the posterior distribution, the variance-covariance matrix is evaluated at the expected outcome, \( y_i = P\{ U(x_i) \} \), for the current parameter value, \( \theta_o \). Since the maximum likelihood estimates are asymptotically normal with mean \( \theta_o \) and variance \( \Sigma_o \), and the prior distribution is normal with mean \( \theta_\pi \) and variance \( \Sigma_\pi \), the posterior distribution, \( \pi(\theta|x) \), is approximately normal with mean \( \theta_p \) and variance \( \Sigma_p \) where

\[
\Sigma_p = (\Sigma_\pi^{-1} + \hat{\Sigma}_o^{-1})^{-1} \quad \text{and} \quad \theta_p = \Sigma_p(\Sigma_\pi^{-1}\theta_\pi + \Sigma_o^{-1}\theta_o).
\]
The variance–covariance matrix, $\hat{\Sigma}$, evaluated at the maximum likelihood estimate $\hat{\theta}$ is substituted in for $\Sigma_o$ since they are asymptotically the same. A derivation of the posterior distribution is included in appendix B.

The multi–stage design for an integration point, $\theta_s$, is constructed one stage at a time using the methods described above. The criterion for this design is calculated using all of the doses and subjects from the individual stages. This entire procedure is repeated for each integration point in the Hermitian formulae to arrive at the expected criterion for the multi–stage design with prior distribution, $\pi(\theta)$.

An alternative method of calculating the criterion for a multi–stage design is to substitute the posterior variance–covariance matrix from the last stage for the asymptotic variance–covariance matrix.

3.4 Multi–Stage Simulation Procedure

Some asymptotic assumptions are made to arrive at the multi–stage design, as discussed in the previous section. Many investigations in cancer research are conducted with small or moderate sample sizes for reasons of cost and the availability of subjects. A simulation study was conducted using the methods discussed in this section to determine how sensitive these designs are to the assumption of large sample sizes.

In the simulation study, one thousand parameter values are drawn randomly from the prior distribution. A multi–stage design is obtained for each set of parameter values. The design for the first stage is fixed by the prior distribution since there is no further prior information. Given a simulated parameter value, $\theta_s$, data is simulated for each stage to obtain the multi–stage design.

The data is simulated at each stage using the parameter value, $\theta_s$; the information from this data is used to update the prior distribution. First, the maximum likelihood estimates, $\hat{\theta}_s$, of the parameter values are obtained for the simulated data. Since the maximum likelihood estimates are asymptotically normal with mean $\theta_s$ and variance $\Sigma_s$, and the prior distribution is normal with mean $\theta_\pi$ and variance $\Sigma_\pi$, the post-
rior distribution $\pi(\theta|x)$ is approximately normal with mean $\hat{\theta}_p$ and variance $\Sigma_p$ (see equation 3.3). The maximum likelihood estimates $\hat{\theta}_s$ and the associated variance-covariance matrix $\hat{\Sigma}_s$ are substituted for $\theta_s$ and $\Sigma_s$ since they are asymptotically the same.

The multi-stage design is constructed for the simulated parameter value, $\theta_s$, using the methods described above. Maximum likelihood estimates of $\theta_s$ and the associated asymptotic variance-covariance matrix are obtained for the multi-stage design using all of the doses and subjects. The distribution of the estimate of the quantity of interest (such as the $ED_{50}$) is found using propagation of error; it is compared to the asymptotic distribution to determine the effects of small sample sizes.
Chapter 4

Single-Stage Design

4.1 Advantages of a Planned Design

Some researchers argue that applying a uniform design over a reasonable range of doses is sufficient for a dose-response study. They question whether the efficiency of a planned design is worth the additional time and effort [12]. For studies which have either ethical or economical considerations, a planned design is worthwhile if the number of observations necessary to attain the desired precision in estimation is decreased.

Müller [10] shows that when constrained to considering only designs with an equal subject allocation to the dosage levels, the uniformly spaced design becomes more efficient as doses are added. The optimal uniform design has as many dosage levels as there are subjects. The complexity of this design makes it impractical. A uniform design with six dose levels is more practical and attains efficiency comparable to that of the optimal uniform design.

For the examples described in section 3.1.3, a uniform design with 6 dose groups equally spaced between the $ED_{05}$ and the $ED_{95}$ is compared to the single-stage design optimized with respect to minimum asymptotic standard deviation. This corresponds, approximately, to doses ranging from -2.5 to 2.5 for the logistic model with a linear response function and from -1.75 to 1.75 for the quadratic response function. The subjects are allocated equally to the dosage levels.

The asymptotic standard deviation of the estimates of the quantiles of response and the slope of the linear response function obtained using the single-stage designs
are compared to those from the uniform design in table 4.1. The parameters are assumed to be \((a, b) = (0, 1)\). This assumption is made without loss of generality since these values can be achieved by shifting and scaling the dose units.

### Advantages of a Planned Design

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Uniform ASD</th>
<th>Single-Stage ASD</th>
<th>Ratio of AV</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ED_{50})</td>
<td>0.257</td>
<td>0.200</td>
<td>1.65</td>
</tr>
<tr>
<td>(ED_{10})</td>
<td>0.490</td>
<td>0.333</td>
<td>2.16</td>
</tr>
<tr>
<td>(ED_{05})</td>
<td>0.615</td>
<td>0.444</td>
<td>1.92</td>
</tr>
<tr>
<td>(b)</td>
<td>0.190</td>
<td>0.151</td>
<td>1.58</td>
</tr>
</tbody>
</table>

**Table 4.1** The asymptotic standard deviation (ASD) of each estimate for the single-stage design is compared to that of the uniform design for the linear response function examples (1 and 2a) in section 3.1.3. Column 4 is the ratio of the asymptotic variance (AV) of the estimate from the uniform design to that of the single-stage design.

The asymptotic variance of each estimate is inversely proportional to the sample size, so the ratio of the asymptotic variance of the estimate from the uniform design to that of the single-stage design is equivalent to the ratio of the sample sizes necessary to obtain the same precision in estimation. The ratio of the asymptotic variances, displayed in column 4, is the increase in sample size needed for the estimate from the uniform design to attain the same precision as that from the single-stage design.

The single-stage design for each of the estimates is significantly more efficient than the uniform design. The increase in sample size necessary for the uniform design to attain the precision of the single-stage design ranges from 58% when estimating \(b\) to 116% when estimating the \(ED_{10}\). This corresponds to efficiencies of the uniform design compared to the single-stage design ranging from 0.63 to 0.46. Central quantiles are much easier to estimate than extreme quantiles. Still, the single-stage design for
estimating the $ED_{50}$ is much better than the uniform design which needs 1.65 times as many subjects for the same precision in estimation.

A similar comparison is made between the uniform design and the planned single-stage design for examples 2b and 3 from section 3.1.3. The first example is designs for estimating the difference between the median doses for two groups, $d_o - d'_o$. The response function is linear and the parameters are $(a, b) = (0, 1)$ and $(a', b') = (-1, 1)$ for groups 1 and 2 respectively. These parameters yield the median doses, $d_o = 0$ and $d'_o = 1$. The other example is designs for estimating $c$, the dose resulting in the maximum response for the quadratic response function. The parameters for this example are $(a, b, c) = (0, 1, 0)$.

Advantages of a Planned Design

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Single-Stage ASD</th>
<th>Uniform ASD</th>
<th>Ratio of AV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_o - d'_o$</td>
<td>0.283</td>
<td>0.368</td>
<td>1.69</td>
</tr>
<tr>
<td>$c$</td>
<td>0.107</td>
<td>0.143</td>
<td>1.79</td>
</tr>
</tbody>
</table>

**Table 4.2** The asymptotic standard deviation (ASD) of each estimate for the single-stage design is compared to that of the uniform design for examples 2b and 3 in section 3.1.3. Column 4 is the ratio of the asymptotic variance (AV) of the estimate from the uniform design to that of the single-stage design.

As seen in table 4.2, the single-stage design has a substantial gain over the uniform design in both examples. A large increase in the number of subjects is required for a uniform design to provide the same precision as the single-stage design. The increase in sample size for the uniform design, ranging from 58% to over 100% in tables 4.1 and 4.2, is substantial for a cancer investigation or other experiment where conditions (such as ethical or economical) require obtaining as much information as possible
from each observation. Good designs can significantly reduce the number of subjects needed to attain the desired precision.

4.2 Designs for Known Parameters

Much of the work on optimal designs in dose–response studies is based on the assumption that the parameters are known \emph{a priori}. Assuming the parameters are known is unrealistic, but it simplifies the optimization procedure; it even allows the optimal design to be solved for analytically in simple cases. Table 4.3 displays the single–stage designs with minimum asymptotic standard deviation for known parameters obtained using the numerical procedure described in section 3.2 on examples 1 and 2a from section 3.1.3; the parameters of the linear response function are \((a, b) = (0, 1)\).

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ED_{50})</td>
<td>50/50</td>
<td>(ED_{48.5}/ED_{51.5})</td>
</tr>
<tr>
<td>(ED_{10})</td>
<td>50/50</td>
<td>(ED_{9}/ED_{11})</td>
</tr>
<tr>
<td>(ED_{0.05})</td>
<td>90.7/9.3</td>
<td>(ED_{8.3}/ED_{91.7})</td>
</tr>
<tr>
<td>(b)</td>
<td>50/50</td>
<td>(ED_{8.3}/ED_{91.7})</td>
</tr>
</tbody>
</table>

\textbf{Table 4.3} The single–stage designs displayed correspond to minimum asymptotic standard deviation of the estimates for examples 1 and 2a in section 3.1.3.

The designs for estimating the quantiles of response are nearly identical to the analytical results discussed in section 2.3. Tosh and McLeish [15] showed that for central quantiles (those falling between the \(ED_{8.3}\) and the \(ED_{91.7}\) for the linear logistic model), the optimal design places all of the subjects at the central quantile. The designs for estimating the central quantiles, \(ED_{50}\) and \(ED_{10}\), found using the numerical optimization procedure have all of the subjects near the quantile. Singularity
problems prevent placement of all the subjects at the central quantile; two dosage levels are needed to fit the linear response function.

The design for estimating the extreme quantile, $ED_{05}$, is identical to the analytically optimal design; allocating the subjects $90.7/9.3$ between $r_1$ and $r_2$. The separation of the optimal designs for estimating quantiles of response into those for central and extreme quantiles (see section 2.3.1) is handled effectively by the numerical algorithm.

The design for estimating the slope of the linear response function places half of the subjects at $r_1$ and the other half at $r_2$. This result ties in nicely with the work by Tosh and McLeish, although they did not look for the optimal design for estimating the slope.

Table 4.4 displays the designs for the other two examples from section 3.1.3. The single-stage design for estimating $d_o - d'_o$ is based on the two-sample logistic model with a linear response function; the parameters are $(b, d_o) = (1, 0)$ and $(b', d'_o) = (1, 1)$. The design for estimating $c$ is for the logistic model with a quadratic response function with parameters $(a, b, c) = (0, 1, 0)$.

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_o - d'_o$</td>
<td>50/50</td>
<td>$d_o/d'<em>o$ lower $ED</em>{18}$</td>
</tr>
<tr>
<td>$c$</td>
<td>50/50</td>
<td>upper $ED_{18}$</td>
</tr>
</tbody>
</table>

Table 4.4 The single-stage designs displayed correspond to minimum asymptotic standard deviation of the estimates for examples 2b and 3 in section 3.1.3.

When trying to detect a difference between the two groups, $d_o - d'_o$, the design with minimum asymptotic standard deviation has the subjects for each group at their
respective $d_o$. This design is intuitive as both $d_o$ and $d'_o$ are central quantiles for their group.

The optimal design for estimating $c$, the dose yielding the maximum response of a quadratic logistic model, contains two doses centered around $c$. This design is not practical since three doses are necessary to fit a quadratic. A variation of this design which is a very good design is to divide the subjects at one of the doses equally between two slightly separated doses. Centering these doses around the original dose levels gives a design close to optimal.

If the parameters are assumed to be known, an intelligently planned single-stage design can be much more efficient than a uniform design (see section 4.1). Since in practice, the true parameter values are not known in advance, some "best guess" of the parameters must be used. A design based on knowledge of the parameter values can be extremely sensitive to misjudgements of the parameter values. Small deviations of the parameter values from the truth can result in a large loss of efficiency.

The effects of small deviations from the true parameter values when estimating the $ED_{50}$ for the linear logistic model are displayed in figure 4.1. The single-stage design for estimating the $ED_{50}$, found using the numerical algorithm in section 3.2, has 50% of the subjects at $d_1 = ED_{48.5}$ and 50% of the subjects at $d_2 = ED_{51.5}$. The asymptotic standard deviation of the estimate of the $ED_{50}$ using this design is calculated to be 0.2001 for a sample size of 100. This is nearly identical to the asymptotic standard deviation of the estimate using the analytically optimal design, 0.2.

The asymptotic variance of the estimate of the $ED_{50}$ using this design is compared with those resulting if $d_1$ and $d_2$ are centered around other quantiles. For example, if $d_1$ and $d_2$ are really centered around the $ED_{46}$, the ratio of the asymptotic variance is approximately six, indicating that about six times as many subjects would be needed to estimate the $ED_{50}$ with the same precision as a correctly placed design. Even an extremely small deviation from the median can result in a large loss of precision. For
Figure 4.1 The ratio of the asymptotic variance of the design for estimating the $ED_{50}$ centered on other quantiles to that of the design centered on the $ED_{50}$ is plotted against the quantile the design was incorrectly centered on.

example, centering $d_1$ and $d_2$ around the $ED_{49}$ has an asymptotic variance ratio of nearly 1.5, so 1.5 times as many subjects are needed to obtain the same precision.

In experimental circumstances, parameter values are not well known in advance. Optimality predicated on prior knowledge of the parameter values is not practical. More robust designs must be obtained, even if the cost of obtaining such designs is some loss of precision. One technique for obtaining more robust designs is to consider
uncertainty in the parameter values as modeled by a prior distribution which describes the uncertainty of the investigator of the parameter values.

4.3 Effect of Uncertainty of Parameter Values

Chaloner [3] finds that considering the investigator's uncertainty of the parameter values changes the designs for which the parameters are known. She assumes a uniform prior distribution and averages the criterion over the prior distribution. She uses criteria which perform well in a very general setting, minimizing either the determinant or weighted trace of the asymptotic variance-covariance matrix.

For small amounts of uncertainty of the parameter values the designs are close to the optimal designs if the mean of the prior distribution is known to be truth. The number of dose levels in the designs increases with uncertainty of the parameters. The increase in number of doses occurs for both central and extreme quantiles. When estimating a central quantile, the doses are added in pairs, centered around the central quantile. Doses are added to the design for estimating an extreme quantile in a skewed pattern; most of the doses are in the tail containing the extreme quantile.

Researchers sometimes have results from previous experiments which provide estimates of the true parameter values. If the estimates are obtained using maximum likelihood estimation techniques, they are asymptotically normal in distribution. The investigator can use these results in the form of a normal prior distribution to describe the uncertainty of the parameter values.

A normal prior distribution is used exclusively in calculating designs which consider uncertainty of the parameters. This places more probability on the values of the parameters more likely to be true. The choice of a specific criterion, minimizing the asymptotic standard deviation of the estimate, also differs from Chaloner.

The effect on the designs for known parameters of considering some uncertainty of the parameters by averaging the criterion over the prior distribution is observed.
The uncertainty of each parameter is increased to explore the effects of increasing uncertainty.

For each example, the mean of the prior distribution is fixed and designs are found for different levels of uncertainty of the parameters. This allows determination of the effect of increasing uncertainty. For the examples involving the linear response function, the mean of the prior distribution is assumed to be \((a, b) = (0, 1)\). In the two-sample example with the linear response function, the mean of both slopes is 1, the mean of \(d_o\) is 0 and the mean of \(d'_0\) is 1. For the example with a quadratic response function, the mean of \((a, b, c)\) is \((0, 1, 0)\).

The designs for estimating the \(ED_{50}\) are shown in table 4.5; the prior variance of \(a\) and \(b\) varies from 0.01 to 1.0. For relatively small levels of uncertainty, \(\text{var}(a) = \text{var}(b) = 0.01\), no doses are added to the design, but the two doses are spread substantially wider than the doses in the design for known parameters, \((a, b) = (0, 1)\).

![Table 4.5](image)

Table 4.5 The single-stage designs for estimating the \(ED_{50}\) of the logistic model with a linear response function are displayed for varying uncertainty of the parameters, \(a\) and \(b\).
The effect of increasing uncertainty of the value of $a$ on the design is the opposite of that for $b$. The design doses spread as the prior variance for $a$ is increased. Initially, the design doses narrow with increases of the prior variance for $b$; dosage levels are added to the design for moderate to large increases in the prior variance of $b$. This effect is displayed graphically in figure 4.2. The top graph displays spreading of the design as prior variance of $a$ is increased. The bottom graph displays a narrowing of the design as the prior variance of $b$ is decreased. The design probably narrows because of the increase in probability for a steep slope as the uncertainty of $b$ increases. A narrow design is necessary to estimate the $ED_{50}$ if the slope is steep.

If the prior variance of the two parameters is increased simultaneously, the design doses initially spread. For moderate to large increases in the prior variance, doses are added to the design in groups of two centered around the $ED_{50}$, surrounding the doses from the previous design; this doses in this design continue to spread as the uncertainty increases.

Even small uncertainty of the parameter values has a great effect on the designs for estimating central quantiles of response near $r_1$ or $r_2$. Recall, the optimal design places all the subjects at the quantile estimated. A prior variance as small as 0.01 on $a$ and $b$ result in a design for estimating the central quantile, $ED_{10}$, resembling designs for estimating an extreme quantile when parameters are known. A large number of subjects are near $r_1$ with the remaining subjects near $r_2$. These designs, displayed in table 4.6, are nearly identical to the designs for estimating the extreme quantile, $ED_{05}$, with the same prior variances of the parameters.

A small amount of uncertainty has little effect on the design for estimating an extreme quantile. These designs are very similar to those for estimating central quantiles near $r_1$ or $r_2$ with the same prior variance. The designs for estimating the $ED_{10}$ and the $ED_{05}$, displayed in tables 4.6 and 4.7, are less complex than the designs found by Chaloner; moderate increases in variance do not result in additional dose levels in the design.
Designs for Estimating ED50

b Known

![Graph showing b Known with various variances of a indicating the effect on the proportion of subjects at different log(dose) values.]

a Known

![Graph showing a Known with various variances of b indicating the effect on the proportion of subjects at different log(dose) values.]

**Figure 4.2** The effect of increasing uncertainty of $a$ on the design for estimating the $ED_{50}$ is displayed in the top graph; the bottom graph shows the effect of increasing uncertainty of $b$. 
Designs for Estimating the $ED_{10}$

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>94/6</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>92/8</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
<td>83/17</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1</td>
<td>86/14</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>87/13</td>
</tr>
<tr>
<td>1.0</td>
<td>0.1</td>
<td>82/18</td>
</tr>
<tr>
<td>0.01</td>
<td>1.0</td>
<td>78/22</td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>78/22</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>80/20</td>
</tr>
</tbody>
</table>

Table 4.6 The single-stage designs for estimating the $ED_{10}$ of the logistic model with a linear response function are displayed for varying uncertainty of the parameters, $a$ and $b$.

Whether estimating the $ED_{10}$ or the $ED_{05}$, as the uncertainty of the parameter values increases, a compromise is made between designs for estimating a central quantile and those for an extreme quantile. For example, the designs for estimating a central or extreme quantile close to $r_1$ are nearly identical; most of the observations are placed near $r_1$ with the remaining placed near $r_2$. These doses narrow as the uncertainty of the slope increases. As the uncertainty of the location parameter increases, the dose level near $r_2$ shifts toward $r_1$. A graphical representation of the effect of increasing prior variance when estimating the $ED_{05}$ is displayed in figure 4.3.

Uncertainty of the parameters when estimating the slope of the linear response function results in the design spreading and dose levels being added. The designs for different prior variances on $a$ and $b$ are displayed in table 4.8.

As the prior variance for both $a$ and $b$ is increased, the dose levels of the design widen. As the uncertainty gets still larger, additional doses are added in groups of two
Figure 4.3 The effect of increasing uncertainty of $a$ on the design for estimating the $ED_{05}$ is displayed in the top graph; the bottom graph shows the effect of increasing uncertainty of $b$. 
Designs for Estimating the $ED_{05}$

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>88/12</td>
</tr>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>87/13</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
<td>86/14</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1</td>
<td>82/18</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>82/18</td>
</tr>
<tr>
<td>1.0</td>
<td>0.1</td>
<td>83/17</td>
</tr>
<tr>
<td>0.01</td>
<td>1.0</td>
<td>74/26</td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>75/25</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>76/24</td>
</tr>
</tbody>
</table>

Table 4.7 The single-stage designs for estimating the $ED_{05}$ of the logistic model with a linear response function are displayed for varying uncertainty of the parameters, $a$ and $b$.

centered around the $ED_{50}$. These designs protect against problems associated with either an extremely large or an extremely small slope. Increasing the uncertainty of the value of the location parameter has little effect on the design. The design narrows slightly as the uncertainty increases. This effect is overwhelmed by increases in uncertainty of the value of the slope.

The effect of uncertainty of the values of $b$, $d_o$ and $d''_o$ in estimating the difference, $d_o - d''_o$, for the two groups is displayed in the designs in table 4.9 for different prior variances on these parameters. Adding a small amount of uncertainty resulted in relatively large changes from the single-stage design obtained from known parameters. The dose level associated with group 1 subjects is at $d_o$, but group 2 subjects are shifted from $d''_o$ to the $ED_{42}$.

Increasing uncertainty for any of the parameters resulted in each groups subjects being divided between two doses centered around $d_o$ and $d''_o$. The separation of these
Designs for Estimating $b$

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>50/50</td>
</tr>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>50/50</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
<td>50/50</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1</td>
<td>50/50</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>50/50</td>
</tr>
<tr>
<td>1.0</td>
<td>0.1</td>
<td>47/6/47</td>
</tr>
<tr>
<td>0.01</td>
<td>1.0</td>
<td>50/50</td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>50/50</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>48/4/48</td>
</tr>
</tbody>
</table>

Table 4.8 The single-stage designs for estimating the slope, $b$, of the linear response function are displayed for varying uncertainty of the parameters, $a$ and $b$.

two doses for each group is determined by the parameter with the larger uncertainty. Large prior variance on the median dose for a particular group spread the dose levels for that group considerably. A large prior variance on the median dose for group one resulted in the doses for group two being centered around the $ED_{38}$ instead of $d''_o$.

A small amount of uncertainty of the parameters when estimating the optimal dose for the quadratic response function results in a design with more doses than the design for known parameters. Recall that the design based on known parameters is two doses centered around the maximum of the quadratic. Adding a small amount of uncertainty of the values of $a$, $b$ or $c$ results in a slight spreading of those two doses and placement of about 10% of the subjects at the maximum of the quadratic. Designs for this estimation with different prior variances are displayed in table 4.10.

Increasing uncertainty has a different effect for the different parameters. Increasing the prior variance of $a$ or $c$ slightly widens the two outside doses in the design,
Designs for Estimating $d_o - d'_o$

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>b 0.01  1.0  0.01</td>
<td>100  50/50</td>
<td>$ED_{50}$  $ED_{32}/ED_{71}$</td>
<td>100  50/50</td>
<td>$ED_{42}$  $ED_{16}/ED_{62}$</td>
</tr>
<tr>
<td>b 1.0  1.0  1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b 0.01  0.01  1.0</td>
<td>50/50</td>
<td>$ED_{64}/ED_{54}$</td>
<td>50/50</td>
<td>$ED_{32}/ED_{63}$</td>
</tr>
<tr>
<td>b 0.01  1.0  0.01</td>
<td>50/50</td>
<td>$ED_{25}/ED_{76}$</td>
<td>50/50</td>
<td>$ED_{38}/ED_{39}$</td>
</tr>
<tr>
<td>b 1.0  0.01  0.01</td>
<td>50/50</td>
<td>$ED_{45}/ED_{55}$</td>
<td>50/50</td>
<td>$ED_{38}/ED_{52}$</td>
</tr>
</tbody>
</table>

Table 4.9 The single-stage designs for estimating $d_o - d'_o$, example 2b from section 3.1.3, are displayed for varying uncertainty of the parameters, $d_o$, $d'_o$, and $b$, the slope common to both groups.

while increasing the prior variance of $b$ narrows the two outside doses. Increases in uncertainty for all three parameters causes a widening of the two outside doses. For moderate uncertainty of the parameter values, dose levels are added to the design, in pairs centered around the optimal dose, $c$.

Using the normal prior distribution to describe the uncertainty of the parameter values and the expected value of the asymptotic standard deviation of the estimate as the criterion results in designs very different from the designs for known parameters, even for small levels of uncertainty. Increases in the uncertainty of the parameter values largely affected the width of the design doses. When estimating $b$, $c$ or the $ED_{50}$ with moderate amounts of uncertainty of the parameters, the designs have additional dosage levels.

In general, increases in uncertainty of the location parameter, $a$, results in a spreading of the dose levels in the design. This is a result of increased probability that the location parameter is far from the true value. The response curve is shifted in that direction. Increases in uncertainty of the slope results in a narrowing of the
Designs for Estimating c

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>1.0</td>
</tr>
<tr>
<td>0.01</td>
<td>1.0</td>
<td>0.01</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4.10 The single-stage designs for estimating the optimal dose of the quadratic response function, c are displayed for varying uncertainty of the parameters, a, b and c.

design doses. This narrowing is a result of an increased probability of a steep slope. A wide design coupled with a steep slope can result in an essentially infinite asymptotic variance.

The methods of determining whether to add dose levels to obtain a better design, discussed in section 3.2, can affect the number of dose levels in the final design. When the best design for a given number of doses is used to determine if additional doses improve the design (step 3 in section 3.2), the final designs tend to be much more complex with little improvement in precision. The optimal design with the subjects equally allocated more effectively determines if additional dosage levels significantly improved the design.

The optimization procedure allows only single doses to be added to the current design; this can be a limiting factor in some cases. One example where this problem is encountered is when finding a design for estimating the slope of the linear response function, see table 4.10. For increasing uncertainty of the parameters, the design has additional doses added in pairs centered around the $ED_{50}$. Adding a single dose and optimizing results in a three dose design which is not centered around the $ED_{50}$. 
Even though the four dose design which allocates 48/2/2/48 subjects to the doses $ED_{02}/ED_{42}/ED_{58}/ED_{98}$ is better than any three dose design, attempts to add a fourth dose to the uncentered three dose design do not yield a better four dose design.

4.4 Designs for Maximizing Power

Power is the probability of correctly rejecting the null hypothesis. We look at designs which maximize the power of a test designed to show a difference between two groups. For the limited sample sizes in many cancer investigations there are not enough subjects to detect a moderate change in slope, so the slope is assumed to be the same in each group. The null hypothesis for showing a difference between groups is $H_o : d_o - d'_o = 0$.

Consider the two–sample example with the parameter values known to be $(b, d_o) = (1, 0)$ and $(b', d'_o) = (1, 1)$. The single–stage optimization procedure determines that the design which places all of group one’s subjects at $d_o$ and all of group two’s subjects at $d'_o$ maximizes the power of the test, $H_o : d_o - d'_o = 0$. This design is identical to the design which minimizes the asymptotic standard deviation of the estimate of $d_o - d'_o$.

The effect of the uncertainty on this design is shown in table 4.11 with the prior variance of each parameter varying from 0.01 to 1.0.

Increasing uncertainty results in wider placement of dose levels. These designs are similar to those with minimum asymptotic standard deviation of the estimate of $d_o - d'_o$. When uncertainty of all the parameters is increased, dose levels for both groups widen, but remain centered around their respective $d_o$.

Increasing the uncertainty of $d_o$ or $d'_o$ effects the design which maximizes power in a slightly differently manner than the design which minimizes asymptotic standard deviation of the estimate of $d_o - d'_o$. Not only do the doses for the group with increasing uncertainty of the median dose separate, they also are shifted towards the median dose for the other group. In the designs which minimize the asymptotic standard deviation of the estimate of $d_o - d'_o$, increases in the uncertainty of $d'_o$ has no shifting
Designs for Testing $H_0 : d_o - d'_o$

<table>
<thead>
<tr>
<th>Prior Variance</th>
<th>$d_o$</th>
<th>$d'_o$</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
<th>Subject Allocation</th>
<th>Dosage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>50/50</td>
<td>$ED_{42}/ED_{60}$</td>
<td>50/50</td>
<td>$ED'<em>{40}/ED'</em>{58}$</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>50/50</td>
<td>$ED_{22}/ED_{70}$</td>
<td>50/50</td>
<td>$ED'<em>{21}/ED'</em>{78}$</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>1.0</td>
<td>50/50</td>
<td>$ED_{38}/ED_{62}$</td>
<td>50/50</td>
<td>$ED'<em>{16}/ED'</em>{59}$</td>
</tr>
<tr>
<td>0.01</td>
<td>1.0</td>
<td>0.01</td>
<td>50/50</td>
<td>$ED_{41}/ED_{85}$</td>
<td>50/50</td>
<td>$ED'<em>{38}/ED'</em>{62}$</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
<td>0.01</td>
<td>50/50</td>
<td>$ED_{37}/ED_{63}$</td>
<td>50/50</td>
<td>$ED'<em>{39}/ED'</em>{61}$</td>
</tr>
</tbody>
</table>

**Table 4.11** The single-stage designs that maximize the power of the test $H_0 : d_o - d'_o$ are displayed for varying uncertainty of the parameters, $d_o, d'_o$ and $b$, the slope common to both groups.

effect, and increases in the uncertainty of $d_o$ shifts the doses for group two towards $d_o$. 


Chapter 5

Multi–Stage Design

The multi–stage design uses the results from preceding stages in the study to improve the design for the succeeding stages. Updating the design allows the multi–stage design to perform as well as or better than the single–stage design. If the parameter values are known fairly accurately, the advantages of the multi–stage design are nonexistent, since adapting to chance outcomes can only worsen a design optimal at the true parameter values.

5.1 Stage Sizes in Multi–Stage Designs

A large factor in the efficiency of a multi–stage design is the allocation of subjects between the stages. There is little difference between a multi–stage design with most of the subjects in the first or last stage and a single–stage design.

A study of two–stage designs was conducted to determine the optimal subject allocation and to explore the effects of increasing uncertainty of the parameter values. Changes in uncertainty of the parameter values does effect the optimal subject allocation; the effect is different for different parameters.

5.1.1 Optimal Stage Sizes for Two–Stage Designs

The effect of varying degrees of uncertainty of the location parameter on the optimal allocation of subjects to the stages of a two–stage design for estimating the $ED_{50}$ is displayed in figure 5.1. The asymptotic standard deviation of the estimate is calculated for different allocations of subjects to the first stage of a two–stage design for estimating the $ED_{50}$ with 100 total subjects. Each line in figure 5.1 is a plot for
a specific level of uncertainty of the location parameter of the number of subjects in stage one versus the asymptotic standard deviation of the estimate of the $ED_{50}$ for that design. Each curve has a minimum which corresponds to the optimal subject allocation for stage one.

**Figure 5.1** The number of subjects in stage one of a two-stage design for estimating the $ED_{50}$ is plotted against the asymptotic standard deviation (ASD) of the estimate of the $ED_{50}$ for prior standard deviations of $a$ of 0.25, 0.5 and 1.0. The prior standard deviation of $b$ is 0.25.

The allocation of subjects to the first stage of the two-stage design for estimating the $ED_{50}$ decreases as the uncertainty of the location parameter increases. The optimal subject allocation for a prior standard deviation of 0.25 on the location parameter is 41 subjects in the first stage and 59 in the second stage. For a prior standard deviation of 1.0 on the location parameter, the optimal allocation is 25
subjects in the first stage and 75 in the second stage. For larger uncertainty of the location parameter, preliminary results are evaluated earlier to allow for quick adjustments of the design.

Uncertainty of the location parameter does not have as large an effect on the optimal subject allocation to the two stages when an extreme quantile is estimated as when the $ED_{50}$ is estimated. The results for estimation of the extreme quantile, $ED_{0.05}$, are displayed in figure 5.2. Once again, the slope is assumed to have a prior standard deviation of 0.25.

![Effect of Uncertainty of $a$ on 2-Stage Design for Estimating $ED_{0.05}$](image)

**Figure 5.2** The number of subjects in stage one of a two-stage design for estimating the $ED_{0.05}$ is plotted against the asymptotic standard deviation (ASD) of the estimate of the $ED_{0.05}$ for prior standard deviations of $a$ of 0.25, 0.5 and 1.0. The prior standard deviation of $b$ is 0.25.
The first stage becomes smaller as the uncertainty of the location parameter increases, but not as quickly as when estimating central quantiles. For a prior standard deviation of 0.25 for the location parameter, the optimal allocation is 33 subjects in the first stage and 67 in the second stage. When the prior standard deviation is 1.0 for the location parameter, the optimal allocation is 24 subjects in the first stage and 76 in the second stage.

The optimal allocation of subjects in the two-stage design for estimating central quantiles close to $r_1$ or $r_2$ is similar to that for estimating extreme quantiles. The first stage gets smaller as the uncertainty of the parameter values is increased, but at a smaller rate than for central quantiles.

Figure 5.3 displays the effect of varying degrees of uncertainty of the slope on the optimal split of a two-stage design for estimating the $E_D_{50}$. The effect of varying the uncertainty of the slope on the optimal subject allocation of the two-stage design for estimating the $E_D_{50}$ is not as large as the effect of varying the uncertainty of the location parameter. The first stage does get smaller as uncertainty of the slope gets larger; but at a very gradual rate.

The effect of uncertainty of the slope on the optimal subject allocation to the two stages when estimating the extreme quantile, $E_D_{05}$, is very similar to that when estimating the $E_D_{50}$. These results are displayed in figure 5.2. Increasing the prior standard deviation of the slope from 0.25 to 1.0 causes the optimal subject allocation for stage one to go approximately from 40 to 20.

One can see in figures 5.1 and 5.2 that as the uncertainty of the location parameter is increased, the asymptotic standard deviation of the estimate increases for every possible subject allocation. This makes sense as the increased uncertainty results in a loss in precision in estimating.

Increasing the uncertainty of the slope does not always result in increased asymptotic standard deviation of the estimate, as displayed in figures 5.3 and 5.4. The asymptotic standard deviation initially gets larger as the uncertainty of the slope
Effect of Uncertainty of b on 2-Stage Design for Estimating ED50

Figure 5.3 The number of subjects in stage one of a two-stage design for estimating the $ED_{50}$ is plotted against the asymptotic standard deviation (ASD) of the estimate of the $ED_{50}$ for prior standard deviations of $b$ of 0.25, 0.5 and 1.0. The prior standard deviation of $a$ is 0.25.
Figure 5.4  The number of subjects in stage one of a two-stage design for estimating the $ED_{05}$ is plotted against the asymptotic standard deviation (ASD) of the estimate of the $ED_{05}$ for prior standard deviations of $b$ of 0.25, 0.5 and 1.0. The prior standard deviation of $a$ is 0.25.

increases. When the uncertainty of the slope reaches a certain level, the asymptotic standard deviation of the estimate actually decreases for every possible subject allocation. This is a result of the increased probability of an extremely large slope. Quantiles are much easier to estimate when the slope is large.

One effect of increasing uncertainty, which is seen for both the location parameter and the slope, is that the designs are more sensitive to subject allocation. The loss of precision which results from not using the optimal subject allocation is more severe when uncertainty of either the location parameter or the slope is large.
5.2 Improvement over Single-Stage Design

Most of the improvement of a multi-stage design over a single-stage design is achieved in the first few stages. This is fortunate, since designs with more than two or three stages are too time consuming to be practical in many investigations.

The improvement in efficiency of a multi-stage design over a single-stage design is explored by comparing a two-stage design to a single-stage design. If the two-stage design is significantly better than the multi-stage design, then multi-stage designs in general are better; additional stages will result in further but lesser improvement.

The two-stage designs compared to the single-stage designs with 100 subjects have the subjects allocated 30/70. As displayed in figure 5.1, this is close to the optimal split of subjects for the level of uncertainty to be considered.

The two-stage designs for the linear response function examples in section 3.1.3 are compared to single-stage designs. The prior distribution of the parameters for these designs is

$$
\pi((a, b)^T) \sim N \left([0, 1]^T, I\right).
$$

(5.1)

were $I$ is the identity matrix. The results for the linear response examples are in table 5.1.

The ratio of the asymptotic variance of the single-stage design to the two-stage design for all three of the quantile of response estimates is around 1.5. This indicates that the single-stage design would need 50% more subjects to attain the same precision as the corresponding two-stage design. When estimating the slope, the single-stage design would require 30% more subjects.

Table 5.2 displays the comparisons between two-stage designs and single-stage designs for estimating $d_o - d_o'$ and $c$ (examples 2b and 3 from section 3.1.3). The prior distribution for the parameters in the two-sample example is
Advantages of a Two–Stage Design

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Single–Stage ASD</th>
<th>Two–Stage ASD</th>
<th>Ratio of AV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ED_{50}$</td>
<td>0.249</td>
<td>0.204</td>
<td>1.49</td>
</tr>
<tr>
<td>$ED_{10}$</td>
<td>0.426</td>
<td>0.343</td>
<td>1.54</td>
</tr>
<tr>
<td>$ED_{05}$</td>
<td>0.544</td>
<td>0.434</td>
<td>1.57</td>
</tr>
<tr>
<td>$b$</td>
<td>0.160</td>
<td>0.141</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Table 5.1  The asymptotic standard deviation (ASD) of each estimate for the two–stage design with 100 subjects allocated 30/70 is compared to that of the single–stage design with 100 subjects for the linear response function examples, 1 and 2a, in section 3.1.3. Column 4 is the ratio of the AV (asymptotic variance) of the estimate from the single–stage design to that of the two–stage design. (See equation 5.1 for the prior distribution of the parameters.)

$$\pi([b, d_o, d'_{o}]^T) \sim N \left([1, 0, 1]^T, \mathbf{I}\right).$$  
(5.2)

For the quadratic response function example, the prior distribution of the parameters is

$$\pi([a, b, c]^T) \sim N \left([0, 1, 0]^T, \mathbf{I}\right).$$  
(5.3)

The two–stage design with the largest gain over its associated single–stage design is that for estimating the optimal dose for the quadratic response function, $c$. The single–stage design needs 2.9 times as many subjects to attain the same precision as the two–stage design. For the two–sample example, the single–stage design requires 30% more observations to estimate with the same precision.

Multi–stage designs should be at least as good as single–stage designs, since they allow early results to improve later portions of the design. A comparison of the single–stage design to the multi–stage design over the entire parameter space reveals
Advantages of a Two-Stage Design

<table>
<thead>
<tr>
<th>Quantity Estimated</th>
<th>Single-Stage ASD</th>
<th>Two-Stage ASD</th>
<th>Ratio of AV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_o - d_o''$</td>
<td>0.347</td>
<td>0.310</td>
<td>1.3</td>
</tr>
<tr>
<td>$c$</td>
<td>0.234</td>
<td>0.138</td>
<td>2.9</td>
</tr>
</tbody>
</table>

**Table 5.2** The asymptotic standard deviation (ASD) of each estimate for the two-stage design 100 subjects allocated 30/70 is compared to that of the single-stage design with 100 subjects for the two sample and quadratic response function examples, 2b and 3, in section 3.1.3. Column 4 is the ratio of the AV (asymptotic variance) of the estimate from the single-stage design to that of the two-stage design. (See equations 5.2 and 5.3 for the prior distributions of the parameters.)

the performance of the multi-stage design for all parameter values. This comparison is shown by a contour plot of the ratio of the asymptotic variance of the single-stage design to that of the two-stage design for a range of parameter values.

Figure 5.5, relating to the estimation of the $ED_{50}$, is a contour plot of the ratio of the asymptotic variances of the single-stage design with 100 subjects to the corresponding two-stage design with a 30/70 allocation of subjects to the two stages; the parameters range over 3 standard deviations from the mean in each direction. The regions marked with dots have an asymptotic variance ratio of less than one; the single-stage design is more efficient than the two-stage design in these regions.

This plot indicates that the two-stage design is much better that the single-stage design for most parameter values. There are a few locations where the single-stage design does slightly better than the two-stage design; for example, parameter values surrounding (-4,-1), (-2,4) and (4,-1).

The single-stage design for estimating the $ED_{50}$ when the prior variance on the parameters is $\text{var}(a) = \text{var}(b) = 1$, from table 4.5, divides the subjects in ratios of
Figure 5.5  The ratio of the asymptotic variance of the estimate of the $ED_{50}$ for a single-stage design to that of a two-stage design is displayed for a range of $a$ and $b$. The prior distribution of the parameters is equation 5.1.
\(4/41/43/12\) between the the doses \(ED_{02}/ED_{33}/ED_{61}/ED_{89}\). The actual dose levels for this design, which are based on the parameter values \((a, b) = (0, 1)\), are \((d_1, d_2, d_3, d_4) = (-4.0, -0.71, 0.46, 2.07)\).

For those regions of the parameter space where the single–stage design outperforms the two–stage design, one of the dose levels in the single–stage design is close to the \(ED_{50}\), \(U(d) \approx 0\). For example, \(U(-4.0)\) is approximately 0 for \((a, b) = (-4, -1)\). Similarly, \(U(0.46)\) is nearly 0 for the parameter values \((-2, 4.14)\). In these examples, the single–stage design has enough subjects placed optimally, if those parameter values are correct, to outperform the two–stage design.

The improvement of the two–stage design over the single–stage design for estimating an extreme quantile is displayed in figure 5.6, which plots the ratio of the asymptotic variances when estimating the \(ED_{05}\). Once again, most of the parameter space has a ratio of variances much larger than one.

For the regions of the plot where the contour is less than one, the parameter values have at least one dose level which is close to the \(ED_{05}\), i.e. \(U(d) \approx -2.95\). From table 4.5, the single–stage design has 76 subjects at the \(ED_{24}\) and 24 at the \(ED_{76}\); these correspond to the doses -1.18 and 1.18. The ratio of variances is less than one near parameters \((-4, 1)\) at which \(U(1.18)\) is close to -2.95.

The two–stage design performs better than the single–stage design for almost all parameter values for the other examples as well. In summary, the single–stage design requires 50% more subjects than the two–stage design to attain the same level of precision in estimation. For most parameter values, the two–stage design attains more precise estimates than the single–stage design.

### 5.3 Effects of Small Sample Sizes

The criteria discussed in section 3.1.3, including the power of a test, are based on the asymptotic variance–covariance matrix. The quality of the designs found using these criteria are dependent on the validity of asymptotic approximations. The simulation
Figure 5.6 The ratio of the asymptotic variance of the estimate of the $ED_{05}$ for a single-stage design to that of a two-stage design is displayed for a range of $a$ and $b$. The prior distribution of the parameters is equation 5.1.
study described in section 3.4 studies the effect that small sample sizes have on designs which minimize the asymptotic standard deviation.

For each set of simulated parameters, \( \theta_x \), a multi-stage design is obtained by simulating data for each stage. For each stage, the maximum likelihood estimates of the parameters from the simulated data and the information matrix are used, along with the prior distribution, to calculate the posterior distribution. The posterior distribution is used in place of the prior distribution for finding the design for the next stage.

The effects of small sample sizes on the multi-stage design is observed by using all of the doses and subjects from the multi-stage design associated with each simulated parameter value to calculate maximum likelihood estimates of the parameter values, \( \theta_x \). The asymptotic density of the difference between \( \hat{\theta} \) and the true parameter values is

\[
(\hat{\theta} - \theta) \sim AN \left( 0, \hat{\Sigma} \right),
\]

where \( \hat{\Sigma} \) is the variance-covariance matrix of \( \hat{\theta} \). A density estimate is obtained of the difference between the quantity estimated and its maximum likelihood estimate. For example, if the quantity of interest is \( ED_{50} \), a density estimate is obtained of the difference between \( \hat{ED}_{50} \) and the simulated values of \( ED_{50} \). This density estimate is compared to the asymptotic density, \( f_{(ED_{50} - ED_{50})} \); a large difference suggests that asymptotic assumptions are largely effected by small sample sizes.

The shape of the density estimate of the difference between the quantity of interest obtained from the the simulated parameter values and the maximum likelihood estimates should be similar to that of the asymptotic density if the sample size for the simulation is sufficient for the asymptotic assumptions to be valid.

This asymptotic density of the difference between the quantity of interest obtained from the simulated parameter values and its maximum likelihood estimate is calculated using \( f_{(\hat{\theta} - \theta)} \) and propagation of error. Calculating \( f_{(\hat{\theta} - \theta)} \) is relatively simple for
known parameter values. The asymptotic density of the difference between \( \hat{\theta} \) and the true parameter values for designs from the criterion averaged over a prior distribution is the asymptotic density, \( f(\hat{\theta} - \theta) \), averaged over the prior distribution, so

\[
\tilde{f}(\hat{\theta} - \theta) = \int \Theta f(\hat{\theta} - \theta) \pi(\theta) d\theta.
\]

The asymptotic density, \( \tilde{f}(\hat{\theta} - \theta) \), is approximated numerically using Hermitian integration in a manner similar to that discussed in appendix A. For each integration point, the maximum likelihood estimate of the parameters, \( \hat{\theta} \), is obtained for the final multi-stage design. The density, \( f(\hat{\theta} - \theta) \), is calculated using a normal distribution with \( \hat{\Sigma} \) as the variance-covariance matrix.

To observe the small sample effect, two-stage designs with 30% of the subjects in stage one and 70% in stage 2 are obtained for sample sizes of 25, 50 and 100. These designs are for the normal prior distribution

\[
\pi \left( \begin{bmatrix} a \\ b \end{bmatrix} \right) \sim N \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).
\]

One thousand parameters are simulated for each design.

Figures 5.7-5.8 contain density estimates of the difference between the quantile of response obtained from the simulated parameter values and its maximum likelihood estimate with a numerical approximation of the corresponding asymptotic density overlaid. The density estimate used in these graphs is the average shifted histogram.

Figure 5.7 is density estimates of (\( \hat{E}D_{50} - ED_{50} \)) for the two-stage design for estimating the \( ED_{50} \) with the asymptotic density overlaid. These figures represent outcomes of trials with sample sizes 25 and 100. The distribution of the maximum likelihood estimates of \( ED_{50} \) is close to the asymptotic distribution, even for sample sizes as small as 25. The maximum likelihood estimated have a slightly smaller variance than expected for sample size 25.

Similarly, density estimates of (\( \hat{E}D_{10} - ED_{10} \)) and (\( \hat{E}D_{05} - ED_{05} \)) are displayed in figures 5.8 and 5.9 respectively. For sample size 100, the density estimates are
Asymptotic Density of MLE's for Estimating ED50

Sample Size 100

Sample Size 25

Figure 5.7 The asymptotic density of \((\hat{E}D_{50} - ED_{50})\) from a two-stage design for estimating the \(ED_{50}\) is overlaid on an average shifted histogram of the 1000 simulated \((\hat{E}D_{50} - ED_{50})\).
close to the asymptotic density. The difference between the density estimates and the asymptotic density is more noticeable for sample size 25, especially in the example for estimating $ED_{05}$. Overall, the effect of small sample sizes on the multi-stage results is acceptable for planning purposes.
Asymptotic Density of MLE's for Estimating ED10

Sample Size 100

ED10.hat-ED10

Sample Size 25

ED10.hat-ED10

Figure 5.8 The asymptotic density of \((\hat{ED}_{10} - ED_{10})\) from a two-stage design for estimating the \(ED_{10}\) is overlaid on an average shifted histogram of the 1000 simulated \((\hat{ED}_{10} - ED_{10})\).
Asymptotic Density of MLE's for Estimating ED05

Sample Size 100

ED05.hat-ED05

Sample Size 25

ED05.hat-ED05

Figure 5.9 The asymptotic density of \((\hat{ED}_{05} - ED_{05})\) from a two-stage design for estimating the \(ED_{05}\) is overlaid on an average shifted histogram of the 1000 simulated \((\hat{ED}_{05} - ED_{05})\).
Appendix A

Expected Value of Criterion Function

Evaluation of the criterion function, \( c(\theta) \), is straightforward for fixed parameter values and designs. To account for the uncertainty of the parameter values, the expected value of the criterion is used; the criterion is averaged over the prior distribution describing the uncertainty, \( \pi(\theta) \). This calculation requires integrating the product of the criterion and the prior distribution over the parameter space,

\[
E_{\pi}[c(\theta)] = \int_{\Theta} c(\theta)\pi(\theta)d\theta. \tag{A.1}
\]

For models with more than one parameter, the integration is over a multivariate normal distribution, usually with nonzero correlations between the parameters.

The simplifying assumption is made that the uncertainty can be described by a multivariate normal distribution, \( \pi(\theta) \sim N(0, \Sigma) \). Using the spectral decomposition theorem, or the Jordan decomposition theorem, \( \Sigma \) can be written as

\[
\Sigma = \Gamma^T \Lambda \Gamma
\]

where \( \Lambda \) is a diagonal matrix of eigenvalues of \( \Sigma \), and \( \Gamma \) is an orthogonal matrix whose columns are standardized eigenvectors. The principle components transformation is \( \Gamma^T \theta \). Serfling [13] shows that the transformed parameters are distributed

\[
\Gamma^T \theta \sim N(0, \Gamma^T \Sigma \Gamma) = N(0, \Lambda).
\]

The multivariate normal distribution has constant density on ellipses or ellipsoids. The principle components transformation rotates the axis so that the transformed parameters lie along the axes of the ellipsoids. The transformation, \( \psi = \Lambda^{-\frac{1}{2}} \Gamma^T \theta \),
not only rotates the axes, but transforms these ellipsoids into circles. The transformed parameters have a standardized multivariate normal distribution, \( \psi \sim N(0, I) \).

Assuming that \( \pi(\theta) \sim N(0, \Sigma) \) is the prior distribution for \( p \) parameters, equation A.1 can be written

\[
\int_{\Theta} c(\theta)\pi(\theta)\,d\theta = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \int_{\Theta} c(\theta) e^{-\frac{\theta^T\Sigma^{-1}\theta}{2}}
\]

Equation A.2 is analytically intractable; however, transforming the parameters to \( \psi \) allows the expected criterion to be written in a form which can be approximated numerically using Hermitian integration.

\[
\int_{\Theta} c(\theta)\pi(\theta)\,d\theta = \frac{1}{(2\pi)^{\frac{p}{2}}|\Gamma^T\Lambda\Gamma|^{\frac{1}{2}}} \int_{\Psi} c(\Gamma^T\Lambda^{-\frac{1}{2}}\psi) e^{-\frac{\psi^T\psi}{2}}|\Gamma^T\Lambda^{-\frac{1}{2}}|\,d\psi
\]

\[
= \frac{1}{(2\pi)^{\frac{p}{2}}} \int_{\Psi} c(\Gamma^T\Lambda^{-\frac{1}{2}}\psi) e^{-\frac{\psi^T\psi}{2}}\,d\psi
\]

\[
= \frac{1}{(2\pi)^{\frac{p}{2}}} \int_{\psi_1} \int_{\psi_2} \ldots \int_{\psi_p} c(\Gamma^T\Lambda^{-\frac{1}{2}}\psi) e^{-(\psi_1^2+\psi_2^2+\ldots+\psi_p^2)}\,d\psi_1\,d\psi_2\ldots d\psi_p
\]

where \( |\Gamma^T\Lambda^{-\frac{1}{2}}| \) is the Jacobian from the transformation. It is equal to \( |\Gamma^T\Lambda\Gamma|^{\frac{1}{2}} \), so they cancel each other.

To arrive at a formula which can be approximated by Hermitian integration [2],

\[
\int e^{-x^2} f(x)\,dx \approx \sum_{i=1}^{n} w_if(x_i),
\]

the transformation \( x_i = \psi_i^2/2 \) is made. This results in the expected criterion,

\[
\int_{\Theta} c(\theta)\pi(\theta)\,d\theta = \frac{1}{(\pi)^{\frac{p}{2}}} \int_{x_1} \int_{x_2} \ldots \int_{x_p} c(\sqrt{2}\Gamma^T\Lambda^{-\frac{1}{2}}x) e^{-(x_1^2+x_2^2+\ldots+x_p^2)}\,dx_1\,dx_2\ldots dx_p
\]

\[
= \frac{1}{(\pi)^{\frac{p}{2}}} \int_{x_1} e^{-x_1^2} \int_{x_2} e^{-x_2^2} \ldots \int_{x_p} e^{-x_p^2} c(\sqrt{2}\Gamma^T\Lambda^{-\frac{1}{2}}x)\,dx_p\ldots dx_2\,dx_1.
\]

where \( x = (x_1, x_2, \ldots, x_p) \). This form is used to arrive at the numerical approximation

\[
\int_{\Theta} c(\theta)\pi(\theta)\,d\theta \approx \frac{1}{(\pi)^{\frac{p}{2}}} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \ldots \sum_{i_p=1}^{n} w_{i_1} w_{i_2} \ldots w_{i_p} c(\sqrt{2}\Gamma^T\Lambda^{-\frac{1}{2}}x).
\]
where \( n \) is the order of integration. The Hermitian weight factor and abscissas for the \( i^{th} \) integration point are \( w_i \) and \( x_i \) respectively [2].

Although the numerical approximation is calculated in the orthogonal parameter space, \( x \) is transformed back into the original parameter space to evaluate the criterion function.
Appendix B

Posterior Distribution of the Next Stage

Suppose that the prior distribution of the vector $\theta$ is normal with mean $\theta_\pi$ and variance-covariance matrix $\Sigma_\pi$. The likelihood function is asymptotically normal with mean $\theta_l$ and variance-covariance matrix $\Sigma_l$, therefore

$$l(\theta|x) \propto \exp\{(\theta - \theta_l)^T \Sigma_l^{-1} (\theta - \theta_l)\}$$

and

$$\pi(\theta) \propto \exp\{(\theta - \theta_\pi)^T \Sigma_\pi^{-1} (\theta - \theta_\pi)\}$$

where $x$ is the observed data. If follows that the product of the prior distribution and the likelihood distribution is

$$f(\theta|x) \propto \exp\{(\theta - \theta_l)^T \Sigma_l^{-1} (\theta - \theta_l)\} + (\theta - \theta_\pi)^T \Sigma_\pi^{-1} (\theta - \theta_\pi). \quad (B.1)$$

The posterior distribution is

$$\pi(\theta|x) = \frac{\pi(\theta)l(\theta|x)}{\int_\Theta \pi(\theta)l(\theta|x)} = \frac{f(\theta|x)}{\int_\Theta f(\theta|x)}. \quad (B.2)$$

If $f(\theta|x)$ is shown to be proportional to a normal distribution,

$$f(\theta|x) \propto \exp\{(\theta - \theta_p)^T \Sigma_p^{-1} (\theta - \theta_p)\}, \quad (B.3)$$

then since the denominator of equation B.2 is a constant, the posterior distribution is normal.

The posterior distribution will be shown to be normal with mean $\theta_p = \Sigma_p(\Sigma_l^{-1}\theta_l + \Sigma_\pi^{-1}\theta_\pi)$ and variance $\Sigma_p = (\Sigma_l^{-1} + \Sigma_\pi^{-1})^{-1}$. Equation B.3 can be expanded
\[ \pi(\theta|x) \propto \exp\{\theta^T \Sigma_p^{-1} \theta - \theta^T \Sigma_p^{-1} \theta_p - \theta_p^T \Sigma_p^{-1} \theta + \theta_p^T \Sigma_p^{-1} \theta_p\}. \] (B.4)

Substituting for \( \theta_p \) and \( \Sigma_p \), we have that \( \Sigma_p^{-1} \theta_p = \Sigma_i^{-1} \theta_i + \Sigma_\pi^{-1} \theta_\pi \) and \( \theta_p^T \Sigma_p^{-1} = \theta_i^T \Sigma_i^{-1} + \theta_\pi^T \Sigma_\pi^{-1} \). Substituting these equalities into equation B.4, along with some algebraic manipulation, it is easily shown that

\[
\begin{align*}
\pi(\theta|x) &\propto \exp\{\theta^T (\Sigma_i^{-1} + \Sigma_\pi^{-1}) \theta - \theta^T (\Sigma_i^{-1} \theta_i + \Sigma_\pi^{-1} \theta_\pi) - (\theta_i^T \Sigma_i^{-1} + \theta_\pi^T \Sigma_\pi^{-1}) \theta\} \\
&\propto \exp\{\theta^T \Sigma_i^{-1} \theta - \theta^T \Sigma_i^{-1} \theta_i - \theta_i^T \Sigma_i^{-1} \theta + \theta^T \Sigma_\pi^{-1} \theta - \theta^T \Sigma_\pi^{-1} \theta_\pi - \theta_\pi^T \Sigma_\pi^{-1} \theta\}
\end{align*}
\]

It can be shown by completing the squares that

\[ f(\theta|x) \propto \exp\{(\theta - \theta_i)^T \Sigma_i^{-1} (\theta - \theta_i)\} + (\theta - \theta_\pi)^T \Sigma_\pi^{-1} (\theta - \theta_\pi)\].

so equations B.1 and B.3 are equal, and the posterior distribution is

\[ \pi(\theta|x) \sim N(\theta_p, \Sigma_p), \]

where \( \theta_p \) and \( \Sigma_p \) are as described above.
Bibliography


