ABSTRACT

Bayesian Models for Discrete Censored Sampling and Dose Finding

Jessica E. Pruszynski, Ph.D.

Chairperson: John W. Seaman, Jr., Ph.D.

We first consider the problem of discrete censored sampling. Censored binomial data may lead to irregular likelihood functions and problems with statistical inference. We consider a Bayesian approach to inference for censored binomial problems and compare it to non-Bayesian methods. We include examples and a simulation study in which we compare point estimation, interval coverage, and interval width for Bayesian and non-Bayesian methods.

The continual reassessment method (CRM) is a Bayesian design often used in Phase I cancer clinical trials. It models the toxicity response of the patient as a function of administered dose using a model that is updated as data accrues. The CRM does not take into consideration the relationship between the toxicity response and the proportion of the administered drug that is absorbed by targeted tissue. Not accounting for this discrepancy can yield misleading conclusions about the maximum tolerated dose to be used in subsequent Phase II trials. We will examine, through simulation, the effect that disregarding the level of bioavailability has on the performance of the CRM.

Bayesian Models for Discrete Censored Sampling and Dose Finding

by

Jessica E. Pruszynski, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

_____

John W. Seaman, Jr., Ph.D., Chairperson

_____

James D. Stamey, Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

Joe C. Yelderman, Jr., Ph.D.

_____

Dean M. Young, Ph.D.

Accepted by the Graduate School
May 2010

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES vi

LIST OF TABLES ix

ACKNOWLEDGMENTS xi

DEDICATION xii

1 Introduction 1

2 The Interval Censoring Problem for the Binomial, Negative Binomial, and Poisson Distributions 3

2.1 The Censored Binomial Distribution 5

2.2 Regularity of the Censored Likelihood 6

2.3 Statistical Inference in the Frequentist Paradigm 8

2.3.1 Maximum Likelihood Estimation 8

2.3.2 Likelihood Intervals 11

2.4 Statistical Inference in the Bayesian Paradigm 14

2.4.1 Derivation of the Posterior Distribution 15

2.4.2 Posterior Moment Derivation 15

2.4.3 Bayesian Credible Intervals 18

2.5 Mixed Precise and Interval Data 23

2.6 Other Distributions 24

2.6.1 Negative Binomial Distribution 25

2.6.2 Poisson Distribution 27

3 Small Sample Interval Sampling for the Binomial Distribution 32
3.1 Data Generation 32
3.2 Simulation Methodology 35
3.3 Simulation Results 38

4 The Effect of Bioavailability on the Continual Reassessment Method 46
4.1 Introduction to the Continual Reassessment Method 46
4.2 Bioavailability and the CRM 47
4.3 Comparing the One Parameter Logistic Model to the Standard Design 49
4.3.1 The CRM and the Standard Design 49
4.3.2 The One-Parameter Logistic Model 50
4.3.3 Problems with the Two-Parameter Logistic Model 50
4.3.4 Simulating Bioavailability 52
4.3.5 Adjusting for Bioavailability 53
4.3.6 Simulation Results 57
4.4 The CRM Power Model 61
4.4.1 The Power Model 61
4.4.2 Adjusting for Bioavailability 62
4.4.3 Simulation Results 64
4.5 Bioavailability: A Maximum Absorbable Dose 69
4.6 Discussion 71

A Censored Binomial Simulation Results 74

B Bioavailability Simulation Results for the One Parameter Logistic Model 87
B.1 90% Bioavailability 87
B.2 75% Bioavailability 88
B.3 35% Bioavailability 89

C Bioavailability Simulation Results for the One Parameter Power Model 90

C.1 90% Bioavailability 90

C.2 75% Bioavailability 91

C.3 35% Bioavailability 92

D Error Plots Comparing the One Parameter Logistic and Power Models 93

E Results from the Maximum Absorbable Dose Scenario 95

BIBLIOGRAPHY 97

LIST OF FIGURES

2.1 Interval Censored Data in Clinical Trials 4

2.2 Comparing the Censored Likelihood to the Precise Likelihood 7

2.3 Comparing the Three Censoring Scenarios 8

2.4 Likelihood Function for $n = 20$ and $1 \leq X \leq 3$ 10

2.5 The Likelihood Function with $c = 0.15$ 14

2.6 Posterior Distributions for $x \in [0, 2]$ and $x \in [3, 5]$ where $\theta \sim beta(1, 1)$. 17

2.7 Likelihood, Prior and Posterior Distributions for Binomial Examples for $n = 5$ 19

2.8 Likelihood, Prior and Posterior Distributions for Binomial Examples for $n = 20$ 22

2.9 Negative Binomial Likelihood where $x \in [7, 10]$ and $r = 1$ 26

2.10 Likelihood, Prior and Posterior Distributions for Negative Binomial Examples 28

2.11 Poisson Likelihood where $x \in [3, 5]$ 30

2.12 Likelihood, Prior and Posterior Distributions for Poisson Examples . . . 31

3.1 Cells of the Multinomial Distribution for $n = 5$ 33

3.2 Likelihood Functions for $n = 5$ 36

3.3 Simulation Results for $\theta = 0.05$ - Unrestricted Censoring Intervals 39

3.4 Relative Frequency Distribution of $k - j$ when $n = 5$ 40

3.5 Relative Frequency Distribution of $k - j$ when $n = 10$ 41

3.6 Relative Frequency Distribution of $k - j$ when $n = 20$ 42

3.7 Simulation Results for $\theta = 0.05$ - Restricted Censoring Intervals 43

3.8 Simulation Results for $\theta = 0.25$ - Unrestricted Censoring Intervals 44

3.9 Simulation Results for $\theta = 0.25$ - Restricted Censoring Intervals 45

3.10 Posterior Distributions for $n = 20$ and $x \in [1, 4]$ 45

4.1 Prior Credible Intervals for the Probability of Dose Limiting Toxicity . . 51
4.2 Logistic Dose Response Curves Adjusted for Bioavailability 55
4.3 Dose Selection for 99% Bioavailability 57
4.4 Number of Patients Assigned to Each Dose Level for 99% Bioavailability 58
4.5 Dose Selection for 65% Bioavailability 59
4.6 Number of Patients Assigned to Each Dose Level for 65% Bioavailability 60
4.7 Power Dose Response Curves Adjusted for Bioavailability 63
4.8 Dose Selection for 99% Bioavailability 64
4.9 Number of Patients Assigned to Each Dose Level for 99% Bioavailability 65
4.10 Dose Selection for 65% Bioavailability 66
4.11 Number of Patients Assigned to Each Dose Level for 65% Bioavailability 67
4.12 MTD Error Plot for 99% Bioavailability 68
4.13 MTD Error Plot for 65% Bioavailability 68
4.14 Dose Selection with 30 Patients - $\eta \sim N(300, 75)$ 70
4.15 Dose Selection with an Average of 14.79 Patients - $\eta \sim Normal(300, 75)$ 71

A.1 Simulation Results for $\theta = 0.10$ - Unrestricted Censoring Intervals 75
A.2 Simulation Results for $\theta = 0.10$ - Restricted Censoring Intervals 76
A.3 Simulation Results for $\theta = 0.50$ - Unrestricted Censoring Intervals 78
A.4 Simulation Results for $\theta = 0.50$ - Restricted Censoring Intervals 79
A.5 Simulation Results for $\theta = 0.75$ - Unrestricted Censoring Intervals 80
A.6 Simulation Results for $\theta = 0.75$ - Restricted Censoring Intervals 81
A.7 Simulation Results for $\theta = 0.90$ - Unrestricted Censoring Intervals 82
A.8 Simulation Results for $\theta = 0.90$ - Restricted Censoring Intervals 83
A.9 Simulation Results for $\theta = 0.95$ - Unrestricted Censoring Intervals 84
A.10 Simulation Results for $\theta = 0.95$ - Restricted Censoring Intervals 85

B.1 Dose Selection for 90% Bioavailability 87
B.2 Number of Patients Assigned to Each Dose Level for 90% Bioavailability 87
B.3 Dose Selection for 75% Bioavailability 88
B.4 Number of Patients Assigned to Each Dose Level for 75% Bioavailability 88
B.5 Dose Selection for 35% Bioavailability 89
B.6 Number of Patients Assigned to Each Dose Level for 35% Bioavailability 89

C.1 Dose Selection for 90% Bioavailability 90
C.2 Number of Patients Assigned to Each Dose Level for 90% Bioavailability 90
C.3 Dose Selection for 75% Bioavailability 91
C.4 Number of Patients Assigned to Each Dose Level for 75% Bioavailability 91
C.5 Dose Selection for 35% Bioavailability 92
C.6 Number of Patients Assigned to Each Dose Level for 35% Bioavailability 92

D.1 MTD Error Plot - 90% Bioavailability 93
D.2 MTD Error Plot - 75% Bioavailability 93
D.3 MTD Error Plot - 35% Bioavailability 94

E.1 Dose Selection with 30 Patients - $\eta \sim Normal(300, 50)$ 95
E.2 Dose Selection with an Average of 13.89 Patients - $\eta \sim Normal(300, 50)$ 95
E.3 Dose Selection with 30 Patients - $\eta \sim Normal(500, 50)$ 96
E.4 Dose Selection with an Average of 14.45 Patients - $\eta \sim Normal(500, 50)$ 96

LIST OF TABLES

2.1 Prior Distributions for Binomial Examples 17

2.2 Posterior Means for $x \in [1, 3]$ and $n = 5$ 18

2.3 Binomial Credible Intervals for $x \in [1, 3]$ and $n = 5$ 21

2.4 Posterior Means for $x \in [1, 3]$ and $n = 20$ 21

2.5 Binomial Credible Intervals for $x \in [1, 3]$ and $n = 20$ 23

2.6 Patients Reporting Drug Side Effects 23

2.7 Negative Binomial Posterior Means and Credible Intervals 27

2.8 Prior Distributions for Poisson Examples 30

2.9 Poisson Means and Credible Intervals 31

3.1 Multinomial distribution of (j, k) for $n = 5$ and $\theta = 0.25$ 34

4.1 Beta(r,s) Distributions for Bioavailability Coefficients 53

4.2 True Dose Response Scenario for 100% Bioavailability 54

4.3 Expected Value of the MTD Using the Logistic Model 56

4.4 Expected Value of the MTD Using the Power Model 64

A.1 Simulation Results for $\theta = 0.05$ - Unrestricted Censoring Intervals 74

A.2 Simulation Results for $\theta = 0.05$ - Restricted Censoring Intervals 75

A.3 Simulation Results for $\theta = 0.10$ - Unrestricted Censoring Intervals 76

A.4 Simulation Results for $\theta = 0.10$ - Restricted Censoring Intervals 77

A.5 Simulation Results for $\theta = 0.25$ - Unrestricted Censoring Intervals 77

A.6 Simulation Results for $\theta = 0.25$ - Restricted Censoring Intervals 78

A.7 Simulation Results for $\theta = 0.50$ - Unrestricted Censoring Intervals 79

A.8 Simulation Results for $\theta = 0.50$ - Restricted Censoring Intervals 80

A.9 Simulation Results for $\theta = 0.75$ - Unrestricted Censoring Intervals 81

A.10 Simulation Results for $\theta = 0.75$ - Restricted Censoring Intervals 82

A.11 Simulation Results for $\theta = 0.90$ - Unrestricted Censoring Intervals 83

A.12 Simulation Results for $\theta = 0.90$ - Restricted Censoring Intervals 84

A.13 Simulation Results for $\theta = 0.95$ - Unrestricted Censoring Intervals 85

A.14 Simulation Results for $\theta = 0.95$ - Restricted Censoring Intervals 86

ACKNOWLEDGMENTS

To my advisor, Dr. John Seaman, I thank you for your constant encouragement and support during this process. Thank you for always believing that I could finish this dissertation even when I didn't believe it myself.

I would also like to express my eternal gratitude to my parents, Charles and Karletta Pruszynski. I could never have made it to this point without your support. I doubt I would be the person that I am today if not for the sacrifices that you have made over the years. Thank you for making the effort to invest in my education and homeschool me twenty years ago. Without that foundation, I don't think I would be as successful academically as I have been. I truly could not have asked for better parents. I love you both!

To the faculty of the Department of Statistical Science, thank you for the support and opportunities that you have provided over the last four years. Thank you for always taking the time to answer all of my questions and for helping me become the best that I can be.

I would especially like to thank the East Texas Baptist University mathematics faculty who first gave me a love for advanced mathematics: Dr. Steve Capehart, Dr. Kevin Reeves, Dr. Melissa Reeves, and Dr. Marty Warren. Without your encouragement, I doubt I would have made the decision to pursue this degree. I must also extend a special thank you to Dr. Marty Warren for encouraging me to pursue statistics in graduate school and especially for pushing me to consider Baylor University!

Finally, I would like to thank Dr. Anna McGlothlin for her valuable input on the continual reassessment method.

DEDICATION

To Him who is able to do immeasurably more than all we ask or imagine.

Soli Deo Gloria

CHAPTER ONE

Introduction

In this dissertation, we investigate Bayesian models for two different problems. We propose models for interval censored counts and compare our approach with a maximum likelihood method. We then consider the problem of sub-bioavailability and its effect on the performance of the continual reassessment method (CRM), a Bayesian dose finding design.

In Chapters 2 and 3, we consider the problem of discrete censored sampling. When we have precise count data, point and interval estimates are easily obtained. However, when we only have censored count data, the derivation of the point and interval estimates becomes more complex. There are three possibilities for the censored likelihood: right-censored, left-censored, and interval-censored. The only one of these possibilities that yields a regular likelihood is the interval-censored case. Because there are problems in interpreting the estimates associated with the irregular likelihood, we focus most of our work on the interval-censored likelihood.

Focusing mainly on the binomial distribution, we calculate the maximum likelihood estimates and likelihood intervals associated with the censored likelihood. We also derive the posterior and marginal distributions and obtain Bayesian point and interval estimates for the binomial distribution. We also consider several examples using the Poisson and negative binomial distributions.

After deriving these point and interval estimates, we conduct a simulation experiment in order to compare the performance of the frequentist and Bayesian estimates for the interval-censored binomial likelihood. In our simulation, we calculate these estimates and compare their performance across different values of the parameter, different Bernoulli sample sizes, and different censoring interval widths.

In our comparison between the two paradigms, we focus on the bias of the point estimates and the width and coverage of the interval estimates.

In Chapter 4, we investigate how low bioavailability levels can influence the performance of the CRM. It is generally accepted in the literature that the Bayesian CRM outperforms the standard $3 + 3$ method for dose finding. We conduct a simulation experiment to investigate whether that remains reasonable when sub-bioavailability is a known issue. In our comparison between the CRM and the $3 + 3$ method, we specifically consider the proportion of trials where the correct dose is selected and the number of patients assigned to each of the possible doses in a single trial. In addition to comparing the CRM and the $3 + 3$ method, we also conduct a simulation experiment to determine how different models of the CRM perform under low bioavailability conditions.

It has been shown that there is a maximum dose that patients can absorb into the body before the system is completely saturated. We consider how the knowledge of this maximum absorbable dose could affect the performance of the CRM. If the knowledge of this dose can significantly shorten the duration of a Phase I trial, in which the CRM is generally used, then more effort should be made into investigating this prior to the start of the trial.

CHAPTER TWO

The Interval Censoring Problem for the Binomial, Negative Binomial, and Poisson Distributions

In many clinical trials, study participants are generally evaluated for the event of interest at the time of their enrollment in the trial and again at several subsequent scheduled times. If the clinical investigators can only monitor the event of interest at these scheduled times, being able to estimate the event incidence per unit of time becomes a central issue in the trial. For example, in an oncology clinical trial, we might schedule patients for monthly visits in order to determine the amount of tumor shrinkage every month. As long as the patients in the trial do not miss their regularly scheduled appointments, we are able to record any changes in tumor size that took place over the last month.

However, patients missing scheduled appointments is a typical problem in many clinical trials. As a result, interval-censored event-time data will arise in trials where clinical investigators are only able to monitor the event of interest at irregular times. When there is this interval-censored data, it becomes much more difficult to estimate the event incidence in the desired time period.

Interval censoring was a problem in AIDS Clinical Trials Groups Study 181, as discussed in Shardell et al. (2007). This trial was a natural history study of cytomegalovirus (CMV) infection in the population of HIV-infected patients. A laboratory test was scheduled every 12 weeks to determine the onset of CMV shedding in the blood. Several of the trial participants were monitored irregularly due to missed or rescheduled appointments. As a result of these missed appointments, the shedding times were only known within an interval of adjacent months as depicted in Figure 2.1.

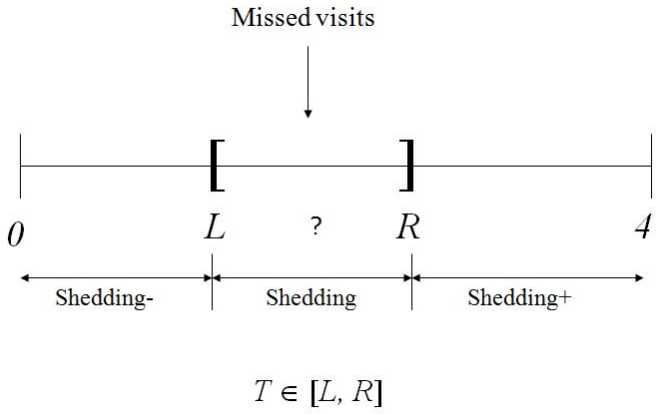

Figure 2.1: Interval Censored Data in Clinical Trials

In this figure, there are three distinct time periods: the time period before shedding occurs, the time period in which shedding occurs, and the time period after shedding occurs. Because the patient missed appointments, an exact time period for the shedding could not be determined. The goal of the study discussed in Shardell et al. (2007) was to estimate the distribution of time to CMV shedding in blood. This is a well-studied problem in survival analysis. See, for example, Klein and Moeschberger (2003).

In this dissertation we are also interested in interval censoring but for counts instead of times. For example, suppose a drug is well known to to produce dizziness as a side effect. In a Phase IV study, patients taking this drug (at a given dose) are asked if they have experienced dizziness in the last week. In the survey instrument, respondents can mark one of "yes", "no", or "not sure" in response to this question. Suppose, in a sample of 100 respondents, we obtain the counts 50, 40, and 10 for "yes", "no", and "not sure," respectively. Then one could model the results as a binomial random variable with a value between 50 and 60. Alternatively, patients may be asked the number of days they experienced dizziness within two hours of taking the drug. Some may be able to answer precisely, such as four days, while

others can only provide an interval, such as 2 to 4 days. How can such information be used to estimate the probability of this adverse event? Chapters 2 and 3 of this dissertation concern problems of this kind.

2.1 The Censored Binomial Distribution

Suppose $x \sim binomial\,(n, \theta)$ where n is the Bernoulli sample size and θ is the parameter we are interested in estimating. The likelihood function for this distribution, given x, is

$$L\,(\theta|n, x) = \binom{n}{x} \theta^x\,(1-\theta)^{n-x}\,.$$

If we observe precise values of x, point and interval estimates can be easily obtained using familiar calculations.

Now suppose that we know the value of the Bernoulli sample size, n, but only know that x falls in the interval $[j, k]$, where, with probability one, $j \geq 0$ and $k \leq n$. That is, instead of observing x directly, we can only observe j and k. Thus, the likelihood becomes

$$L\,(\theta|n, j, k) = \sum_{x=j}^{k} \binom{n}{x} \theta^x\,(1-\theta)^{n-x}\,. \tag{2.1}$$

The likelihood in (2.1) should not be confused with that of a binomial sample. Here the data point consists of j and k and corresponds to an unseen binomial count from a Bernoulli sample of size n.

Pawitan (2001) and Frey and Marrero (2008) consider the censored binomial count problem. Both take a maximum likelihood approach. In this chapter and the next, we consider both maximum likelihood and Bayesian solutions. In this chapter, we develop the basic models and study their properties. In Section 2.2, we study the regularity of the censored likelihood. In Section 2.3, we consider statistical inference in the frequentist paradigm and derive the maximum likelihood estimator and the likelihood intervals. In Section 2.4, we consider statistical inference in the Bayesian paradigm and derive the posterior and marginal distributions, the posterior

moments, and the credible intervals. In Section 2.5, we look at point and interval estimates for the censored negative binomial and Poisson distributions. In Chapter 3, we conduct simulation experiments to study the behavior of the censored binomial model.

2.2 Regularity of the Censored Likelihood

There are three different types of likelihoods we can potentially encounter when working with censored binomial counts. There can either be an upper bound where we know $x \leq k$, a lower bound where we know $x \geq j$, or an interval where we know $j \leq x \leq k$, each with probability one. We first consider the case of the upper bound, and compare its likelihood function to a likelihood function where the precise value of x is known.

Pawitan (2001) cites an example where 100 seeds were planted, and it is known only that $x \leq 10$ seeds germinated; that is we do not know the exact number of seeds that germinated. From (2.1) the likelihood is

$$\begin{aligned} L\left(\theta|100,0,10\right) &= P\left(x \leq 10\right) \\ &= \sum_{x=0}^{10} \binom{100}{x} \theta^x \left(1-\theta\right)^{100-x}. \end{aligned}$$

Suppose another 100 seeds are planted, and we know that exactly 5 seeds germinated. In this case the likelihood is

$$\begin{aligned} L\left(\theta|x=5\right) &= P(x=5) \\ &= \binom{100}{5} \theta^5 (1-\theta)^{100-5}. \end{aligned}$$

We compare the two likelihood functions in Figure 2.2. As Pawitan notes (2001, p. 40) the case for $x < 11$ is not regular and leads to difficulties in the interpretation of pure likelihood intervals. The exact case poses no such problems.

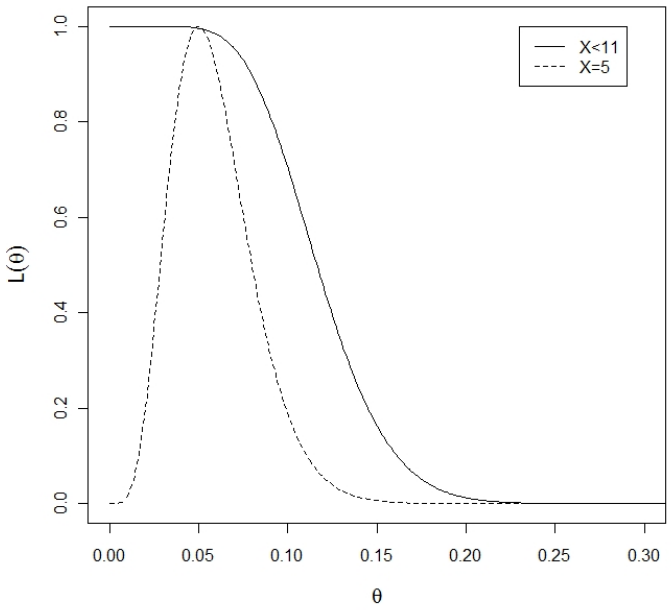

Figure 2.2: Comparing the Censored Likelihood to the Precise Likelihood

As a further illustration, suppose we have a Bernoulli sample of $n = 5$. In Figure 2.3 we compare the likelihoods for observing $x \leq 2$, $x \geq 3$, and $1 < x < 4$. Note that the first two are not regular, and the last is.

In general, when the censoring interval includes 0, the likelihood for θ will be decreasing, and when it contains n, the likelihood will be decreasing. This is evident from (2.1) since for fixed θ, the function $L(\theta|n, j, k)$ is the probability that a random variable, x, distributed $binomial(n, \theta)$, is in the interval $[j, k]$. Viewed in this way, $P(0 \leq x \leq k|\theta)$ decreases as θ increases for fixed $k < n$. Similarly, for fixed $j > 0$, $P(j \leq x \leq n|\theta)$ will increase as θ increases.

In what follows, we will often assume that the observed value, x, is such that $0 < j \leq x \leq k < n$, as it is only in that case that the likelihood analysis yields reasonable interval estimates; i.e. interval estimates that admit an interpretation from the frequentist point of view. As we shall see, the Bayesian analysis is not constrained in this way.

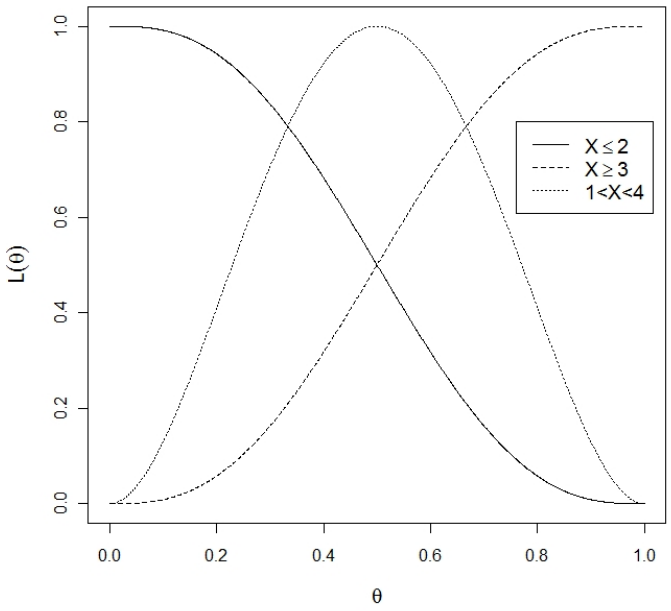

Figure 2.3: Comparing the Three Censoring Scenarios

2.3 Statistical Inference in the Frequentist Paradigm

We now consider the derivation of both point and interval estimates from the frequentist perspective. We begin with the maximum likelihood estimator derived in Frey and Marrero (2008).

2.3.1 Maximum Likelihood Estimation

We consider the likelihood function in (2.1). In the case where $j = 0$ and $k < n$, then the MLE, $\hat{\theta}$, is 0 since it is the only solution that satisfies $L(\hat{\theta}|n, j, k) = 1$. Alternatively, if $j > 0$ and $k = n$, then $\hat{\theta} = 1$. Therefore, Frey and Marrero (2008) focus on the more interesting case where $0 < j < k < n$.

Using the well-known relationship between the binomial and beta distributions (David and Nagaraja, 2003), Frey and Marrero (2008) proceed as follows. We can

rewrite the likelihood as

$$\begin{aligned} L\left(\theta|n,j,k\right) &= I_p(j,n-j+1) - I_p(k+1,n-k) \\ &= \int_0^\theta \frac{n!}{(j-1)!(n-j)!} y^{j-1}(1-y)^{n-j}dy \\ &\quad - \int_0^\theta \frac{n!}{k!(n-k-1)!} y^{k}(1-y)^{n-k-1}, \end{aligned}$$

where I_p represents the incomplete beta function,

$$I_p(a,b) = \frac{1}{B(a,b)} \int_0^p t^{a-1}(1-t)^{b-1}dt,$$

and where $B(a,b)$ is the beta function.

Note that the likelihood above is the difference of two incomplete beta functions. Differentiating, we obtain

$$\frac{dL}{d\theta} = \frac{n!}{(j-1)!(n-j!)}\theta^{j-1}(1-\theta)^{n-j} - \frac{n!}{k!(n-k-1!)}\theta^{k}(1-\theta)^{n-k-1}.$$

After setting $\frac{dL}{d\theta} = 0$ and factoring the derivative, we find that $\hat{\theta}$ must satisfy

$$n!\hat{\theta}^{j-1}\left(1-\hat{\theta}\right)^{n-k-1}\left[\frac{\left(1-\hat{\theta}\right)^{k-j+1}}{(j-1)!(n-j)!} - \frac{\hat{\theta}^{k-j+1}}{k!(n-k-1)!}\right] = 0. \tag{2.2}$$

There are three possible solutions to (2.2). Either $\hat{\theta} = 0$, $\hat{\theta} = 1$, or $\hat{\theta}$ satisfies

$$\frac{\left(1-\hat{\theta}\right)^{k-j+1}}{(j-1)!(n-j)!} = \frac{\hat{\theta}^{k-j+1}}{k!(n-k-1)!}.$$

Since we have three possible solutions, we consider which one maximizes the likelihood. We find that $\hat{\theta} = 0$ and $\hat{\theta} = 1$ satisfy $L(\hat{\theta}|n,j,k) = 0$ and thus minimize the likelihood. Therefore, $\hat{\theta}$ must satisfy

$$\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)^{k-j+1} = \frac{k!(n-k-1)!}{(j-1)!(n-j)!}.$$

This implies that

$$\begin{aligned}\frac{\hat{\theta}}{1-\hat{\theta}} &= \left(\frac{k!(n-k-1)!}{(j-1)!(n-j)!}\right)^{1/(k-j+1)} \\ &= \left(\frac{j(j+1)\cdots k}{(n-j)(n-j+1)\cdots(n-k)}\right)^{1/(k-j+1)} \\ &= \left\{\left(\frac{j}{n-j}\right)\left(\frac{j+1}{n-j-1}\right)\cdots\left(\frac{k}{n-k}\right)\right\}^{1/(k-j+1)}.\end{aligned}$$

After taking the natural logarithm of both sides, we find that

$$\log\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right) = \frac{1}{k-j+1}\sum_{i=j}^{k}\log\frac{i/n}{1-\frac{i}{n}}. \tag{2.3}$$

We now consider an example. Let $n = 20$ and suppose x is known to be in the interval $[1, 3]$. The likelihood function can be seen in Figure 2.4.

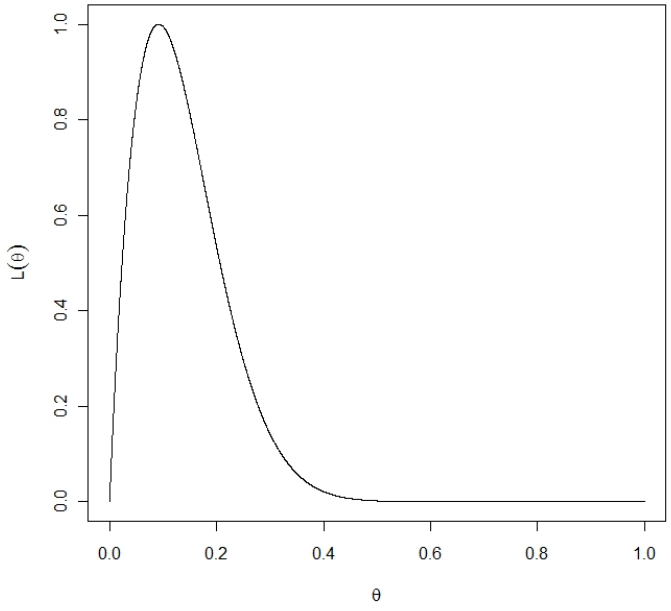

Figure 2.4: Likelihood Function for $n = 20$ and $1 \leq X \leq 3$

Using Equation 2.3,

$$\begin{aligned}\log\frac{\hat{\theta}}{1-\hat{\theta}} &= \frac{1}{3-1+1}\left[\log\frac{1/20}{19/20} + \log\frac{2/20}{18/20} + \log\frac{3/20}{17/20}\right] \\ &= -2.292,\end{aligned}$$

which implies that

$$\begin{aligned}\hat{\theta} &= \frac{\exp(-2.292)}{1+\exp(-2.292)} \\ &= 0.092. \end{aligned} \tag{2.4}$$

2.3.2 Likelihood Intervals

We are often interested in communicating statistical information using only the likelihood function. If this likelihood function is regular, then presenting the MLE and its associated standard error is often sufficient (Pawitan, 2001). However, if the likelihood is not reasonably regular, this may be infeasible since the standard error may not be well defined.

In the cases where the likelihood is not regular, we can construct interval estimates directly from the likelihood function. This method is due to R.A. Fisher - see Pawitan(2001, p. 35) and references therein.

Suppose the random vector $x = (x_1, x_2, \ldots, x_n)$ has independent components with likelihood $L(\theta|x)$ where $\theta = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta$. Let $\hat{\theta}$ be a unique MLE for θ given x. For a fixed constant c, the likelihood interval is defined as

$$\left\{ \theta \in \Theta : \frac{L(\theta|x)}{L(\hat{\theta}|x)} > c \right\}. \tag{2.5}$$

Choosing the cutoff value, c, becomes a central question when utilizing likelihood intervals. In his development of the method, Fisher leaves the selection of the cutoff point open, but does suggest that parameter values with less than 1/15 (or 6.7%) likelihood should be considered suspicious. However, this qualification is not appropriate for every situation. Probabilistic calibration, with a frequentist interpretation, is the most commonly used means of determining the value of c (Pawitan, 2001). This approach requires that the likelihood be approximately regular.

Suppose, for example, the vector of observations, x, constitutes an IID sample from a $N(\theta, \sigma^2)$ distribution. It can be shown that

$$\log \frac{L(\theta|x)}{L(\hat{\theta}|x)} = -\frac{n}{2\sigma^2}(\bar{x} - \theta)^2, \tag{2.6}$$

where $\hat{\theta} = \bar{x}$, the sample mean. We know $\bar{x} \sim N(\theta, \sigma^2/n)$. Therefore,

$$\frac{n}{\sigma^2}(\bar{x} - \theta)^2 \sim \chi^2_1. \tag{2.7}$$

Multiplying both sides of (2.6) by -2 gives us Wilk's likelihood ratio statistic:

$$W \equiv 2 \log \frac{L\left(\hat{\theta}|x\right)}{L(\theta|x)} \sim \chi^2_1. \tag{2.8}$$

The result in (2.8) is, of course, exact for normal sampling. It can be used when sampling from other distributions as long as the corresponding likelihood is approximately regular. This allows a method of confidence interval calibration for a wide range of data models.

Thus, suppose we have an approximately regular likelihood, $L(\theta|x)$, for a possibly vector-valued parameter, θ, and based on a data vector, x. Suppose $\hat{\theta}$ is a unique MLE for θ. We have, for fixed θ,

$$\begin{aligned} P\left(\frac{L(\theta|x)}{L(\hat{\theta}|x)} > c\right) &\approx P\left(2 \log \frac{L(\hat{\theta}|x)}{L(\theta|x)} < -2 \log c\right) \\ &= P\left(\chi^2_1 < -2 \log c\right). \end{aligned} \tag{2.9}$$

Therefore, we can use (2.9) to choose c. For some $0 < \alpha < 1$, we have

$$c = \exp\left(-\frac{1}{2}\chi^2_{1,1-\alpha}\right), \tag{2.10}$$

where $\chi^2_{1,1-\alpha}$ is the $100(1-\alpha)$ percentile of χ^2_1. Therefore, we have

$$\begin{aligned} P\left(\frac{L(\theta|x)}{L(\hat{\theta}|x)} > c\right) &\approx P\left(\chi^2_1 < \chi^2_{1,1-\alpha}\right) \\ &= 1 - \alpha. \end{aligned}$$

By using this method to select c, the likelihood interval is an approximate $100(1-\alpha)\%$ confidence interval for θ. For example, if $\alpha = 0.05$, (2.10) will yield $c = 0.15$.

It is important to note that when we use this methodology to calibrate the likelihood, we are no longer using pure likelihood inference. Because we calibrated the likelihood using the sampling distribution of $\hat{\theta}$, we are now using inference based on the repeated sampling paradigm.

If the likelihood, $L(\theta|x)$ is not regular, then we interpret the interval as a pure likelihood interval. If this is the case, we are unable to calibrate the likelihood and characterize the uncertainty associated with the interval estimate. It is because of this that we will only consider the interval-censored binomial likelihood function when we construct likelihood intervals. The right and left censored binomial likelihoods are clearly irregular, and, for the purposes of the dissertation, we will only consider the case of a regular likelihood when using non-Bayesian methods.

As an illustration, we now construct a likelihood interval for the scenario where $n = 20$ and $x \in [1, 3]$. As we saw in Figure 2.4, the likelihood function appears to be reasonably regular. We are interested in obtaining a 95% interval, so we set $c = 0.15$. This is depicted in Figure 2.5.

The likelihood interval consists of all values of θ that satisfy

$$\frac{L(\theta|20, 1, 3)}{L(\hat{\theta}|20, 1, 3)} > c.$$

From (2.4), we know that $\hat{\theta} = 0.092$, and using that MLE, we find that

$$L\left(\hat{\theta}|n = 20, j = 1, k = 3\right) = 0.749.$$

Therefore, the likelihood interval for this example consists of all values of θ that satisfy

$$\frac{\sum_{x=1}^{3} \frac{20!}{x!(20-x)!}\theta^x \left(1-\theta\right)^{20-x}}{0.749} > 0.15.$$

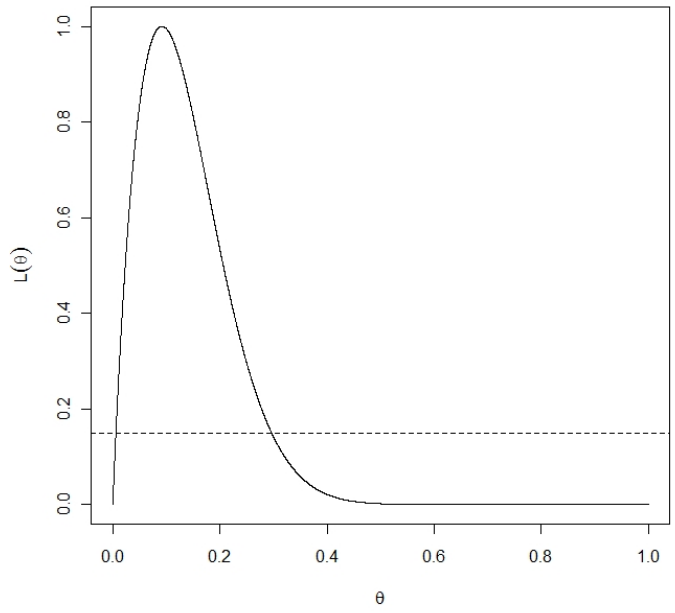

Figure 2.5: The Likelihood Function with $c = 0.15$

Using numerical methods, we find that the likelihood interval is $(0.006, 0.297)$. Since the likelihood function is regular, we can interpret this interval as we would any 95% confidence interval.

2.4 Statistical Inference in the Bayesian Paradigm

We now propose a Bayesian solution to the censored count problem. Consider again the likelihood

$$L(\theta|k) = \sum_{x=j}^{k} \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

As we saw in Figure 2.3, this likelihood is not regular when $j = 0$ or $k = n$. In the former case the MLE is 0 and in the latter it is 1. There is no distribution that will give us a reasonable value for the standard error of this estimator. In addition, we are unable to interpret any likelihood interval estimates as there is no probabilistic foundation for calibrating the likelihood. However, we can easily derive a Bayesian model for this case, and any other censoring interval.

2.4.1 Derivation of the Posterior Distribution

Consider the likelihood function in (2.1) and suppose we let $\theta \sim beta(a, b)$ be our prior distribution. Then, the posterior distribution is

$$\begin{aligned}\pi\left(\theta | j \leq x \leq k\right) &\propto \left[\sum_{x=j}^{k} \binom{n}{x} \theta^{x} \left(1-\theta\right)^{n-x}\right] \frac{1}{B(a,b)} \theta^{a-1} \left(1-\theta\right)^{b-1} \\ &= \frac{1}{B(a,b)} \sum_{x=j}^{k} \binom{n}{x} \theta^{x+a-1} \left(1-\theta\right)^{n-x+b-1}, \end{aligned} \tag{2.11}$$

where $B(a, b)$ is the beta function and both a and b are positive. The marginal distribution is

$$\begin{aligned} m(j,k) &= \int_{0}^{1} \frac{1}{B(a,b)} \sum_{x=j}^{k} \binom{n}{x} \theta^{x+a-1} \left(1-\theta\right)^{n-x+b-1} d\theta \\ &= \frac{1}{B(a,b)} \int_{0}^{1} \sum_{x=j}^{k} \binom{n}{x} \theta^{x+a-1} \left(1-\theta\right)^{n-x+b-1} d\theta \\ &= \frac{1}{B(a,b)} \sum_{x=j}^{k} \binom{n}{x} \int_{0}^{1} \theta^{x+a-1} \left(1-\theta\right)^{n-x+b-1} d\theta \\ &= \frac{1}{B(a,b)} \sum_{x=j}^{k} \binom{n}{x} B(x+a, n-x+b). \end{aligned} \tag{2.12}$$

2.4.2 Posterior Moment Derivation

We now derive the form of the p^{th} moment of the posterior distribution in (2.11). We have

$$\begin{aligned} E\left(\theta^{p} | j \leq x \leq k\right) &= \int_{0}^{1} \frac{\sum_{x=j}^{k} \binom{n}{x} \theta^{p} \theta^{x+a-1} \left(1-\theta\right)^{n-x+b-1}}{\sum_{x=j}^{k} \binom{n}{x} B(x+a, n-x+b)} d\theta \\ &= \int_{0}^{1} \frac{\sum_{x=j}^{k} \binom{n}{x} \theta^{x+a+p-1} \left(1-\theta\right)^{n-x+b-1}}{\sum_{x=j}^{k} \binom{n}{x} B(x+a, n-x+b)} d\theta \\ &= \frac{\sum_{x=j}^{k} \binom{n}{x} B(x+a+p, n-x+b)}{\sum_{x=j}^{k} \binom{n}{x} B(x+a, n-x+b)}. \end{aligned} \tag{2.13}$$

Thus, the posterior mean is

$$E\left(\theta|j \leq x \leq k\right) = \frac{\sum_{x=j}^{k} \binom{n}{x} B(x+a+1, n-x+b)}{\sum_{x=a}^{b} \binom{n}{x} B(x+a, n-x+b)}. \tag{2.14}$$

The posterior variance is

$$\begin{aligned} Var(\theta|j \leq x \leq k) &= E\left(\theta^2|x\right) - \left[E\left(\theta|x\right)\right]^2 \\ &= \frac{\sum_{x=j}^{k} \binom{n}{x} B(x+a+2, n-x+b)}{\sum_{x=a}^{b} \binom{n}{x} B(x+a, n-x+b)} \\ &\quad - \left[\frac{\sum_{x=j}^{k} \binom{n}{x} B(x+a+1, n-x+b)}{\sum_{x=a}^{b} \binom{n}{x} B(x+a, n-x+b)}\right]^2. \end{aligned} \tag{2.15}$$

In our earlier example, where $x \in [1, 3]$ and $n = 20$ the MLE was 0.092. If we use a $beta(1, 1)$ prior for θ, then the posterior mean, from (2.14), is 0.097.

As we discussed in Section 2.3, if $x \in [0, j]$, the MLE will be 0, and when $x \in [k, n]$, the MLE will be 1. We consider two censoring scenarios for our next example: $x \in [0, 2]$ and $x \in [3, 5]$ where $n = 5$. Using a $beta(1, 1)$ prior on θ, we examine the plots of these two posterior distributions in Figure 2.6. We now calculate the posterior means for both of these distributions using (2.14). For $x \in [0, 2]$, we find the posterior mean to be 0.286; for $x \in [3, 5]$, we find the posterior mean to be 0.714. At the least, the Bayesian estimates have the virtue of not being zero or one.

We now consider examples with interval censored likelihoods and more informative priors. We investigate six different prior distributions and consider their effect on the resulting posterior mean. Suppose we have reasonable information that the true value of θ is around 0.25. We set the mean of the prior distribution to be 0.25 and construct three prior distributions with standard deviations 0.05, 0.10, and 0.20. For the second set of prior distributions, we consider the possibility that

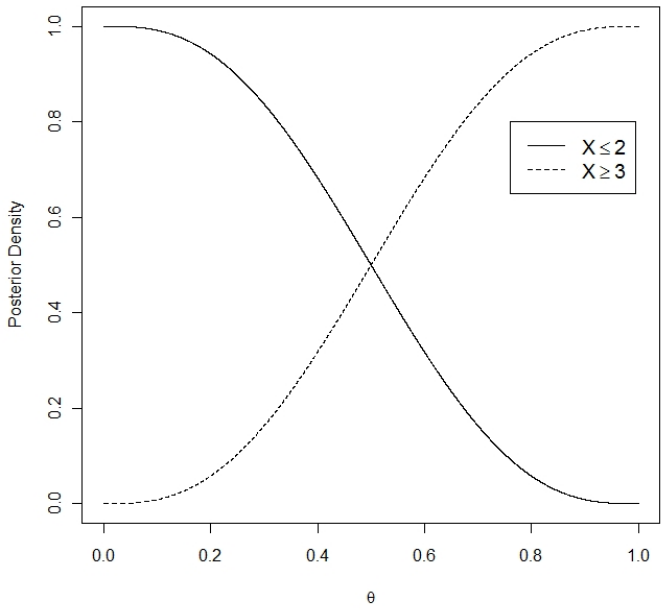

Figure 2.6: Posterior Distributions for $x \in [0, 2]$ and $x \in [3, 5]$ where $\theta \sim beta(1, 1)$

the true value of θ is 0.50 and construct prior distributions using the same standard deviations as in the first three prior distributions. Using these means and standard deviations, we obtain six beta distributions as seen in Table 2.1.

Table 2.1: Prior Distributions for Binomial Examples

Prior Distribution	*Prior Mean*	*Prior SD*
$beta(18.5, 55.5)$	0.25	0.05
$beta(4.4, 13.3)$	0.25	0.10
$beta(0.9, 2.8)$	0.25	0.20
$beta(49.5, 49.5)$	0.50	0.05
$beta(12, 12)$	0.50	0.10
$beta(2.6, 2.6)$	0.50	0.20

In the following examples, we will consider the likelihood when $n = 5$ and $x \in [1, 3]$. Refer to Figure 2.7 for plots of the likelihood, prior distributions, and resulting posterior distributions.

Using Equation 2.14, we calculate the posterior mean for these six distributions displayed in Figure 2.7. The results are displayed in Table 2.2.

Table 2.2: Posterior Means for $x \in [1, 3]$ and $n = 5$

Prior Distribution	*Posterior Mean*
$beta(18.5, 55.5)$	0.254
$beta(4.4, 13.3)$	0.266
$beta(0.9, 2.8)$	0.299
$beta(49.5, 49.5)$	0.497
$beta(12, 12)$	0.488
$beta(2.6, 2.6)$	0.459

These posterior means are as one would expect when using these particular prior distributions. When the standard deviation of the prior is very small, the posterior mean will be very close to the prior mean. However, as the standard deviation of the prior increases, the posterior mean begins to deviate from the prior mean, shrinking toward the MLE.

2.4.3 Bayesian Credible Intervals

We now consider obtaining interval estimates for the censored binomial problem in the Bayesian paradigm. To this end, we construct 95% equal-tailed credible sets. To do this, we must solve the following two integrals for y and z, respectively:

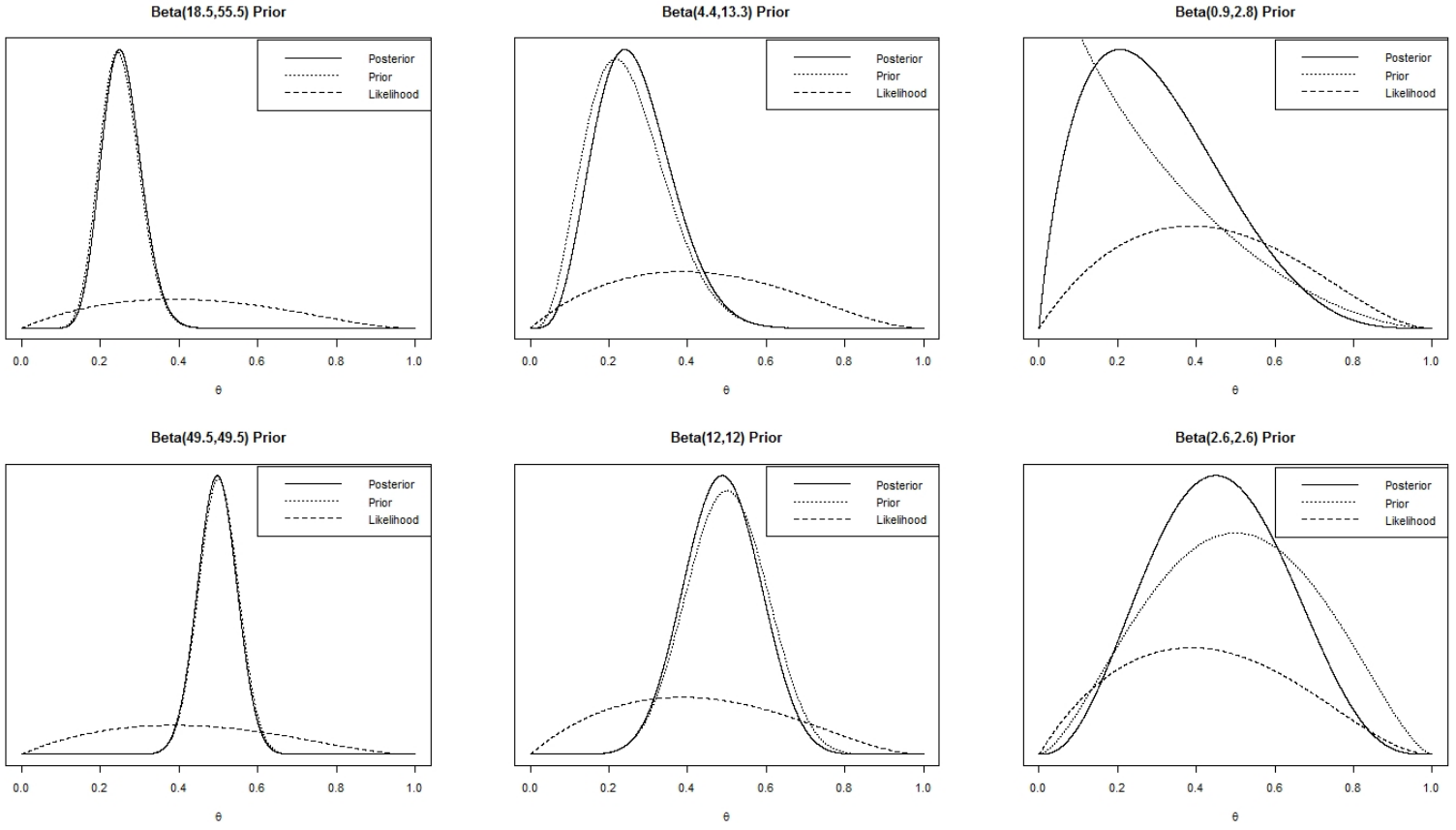

Figure 2.7: Likelihood, Prior and Posterior Distributions for Binomial Examples for $n = 5$

$$\int_0^y \pi(\theta|x)d\theta = 0.025 \tag{2.16}$$

and

$$\int_0^z \pi(\theta|x)d\theta = 0.975. \tag{2.17}$$

For example, suppose again that $x \in [1, 3]$ when $n = 5$. We place a $beta(1, 1)$ prior distribution on θ, and then compare it to the likelihood interval obtained for this same example. Using (2.16) and (2.17), we find the 95% credible interval for θ to be $(0.074, 0.831)$. The true value for θ is contained within this interval with 95% probability. The likelihood interval for θ for this same example is $(0.026, 0.872)$. As we would expect when using a uniform prior distribution, the two intervals do not appear to be significantly different although the Bayesian interval is slightly narrower.

We next consider the censoring interval $x \in [0, 2]$ when $n = 5$. We did not calculate a likelihood interval for this example because there was no probabilistic foundation for calibration and interpretation. A Bayesian analysis with a $beta(1, 1)$ prior yields the credible interval $(0.013, 0.708)$.

Now suppose $x \in [3, 5]$ when $n = 5$. Again using a $beta(1, 1)$ prior distribution, we obtain the credible interval $(0.292, 0.987)$.

We now calculate credible intervals using the prior distributions in Table 2.1 for the case where $x \in [1, 3]$ and $n = 5$. The prior distributions and posterior distributions can be seen in Figure 2.7. The credible intervals are listed in Table 2.3. Because the integrals become increasingly complex with the more informative priors, we are not able to solve the integral directly. Instead, we use numerical methods to solve the integrals and obtain the limits of the credible intervals.

Table 2.3: Binomial Credible Intervals for $x \in [1, 3]$ and $n = 5$

Prior Distribution	*95% Credible Interval*
$beta(18.5, 55.5)$	$(0.164, 0.357)$
$beta(4.4, 13.3)$	$(0.102, 0.473)$
$beta(0.9, 2.8)$	$(0.042, 0.672)$
$beta(49.5, 49.5)$	$(0.4, 0.594)$
$beta(12, 12)$	$(0.305, 0.673)$
$beta(2.6, 2.6)$	$(0.151, 0.781)$

In the previous example, we can see that there was very little updating from the prior distribution due to the relatively flat likelihood. We now consider a likelihood function when $x \in [1, 3]$ and $n = 20$. Using the prior distributions in Table 2.1, we consider the plots of the posteriors in Figure 2.8.

Using Equation 2.14, we calculate the posterior mean for these six distributions displayed in Figure 2.8. The results are displayed in Table 2.4.

Table 2.4: Posterior Means for $x \in [1, 3]$ and $n = 20$

Prior Distribution	*Posterior Mean*
$beta(18.5, 55.5)$	0.223
$beta(4.4, 13.3)$	0.176
$beta(0.9, 2.8)$	0.119
$beta(49.5, 49.5)$	0.439
$beta(12, 12)$	0.333
$beta(2.6, 2.6)$	0.193

Posterior updating is much more in evidence here, in contrast to the $n = 5$ case. Shrinkage toward the MLE can be seen in all six examples. Corresponding

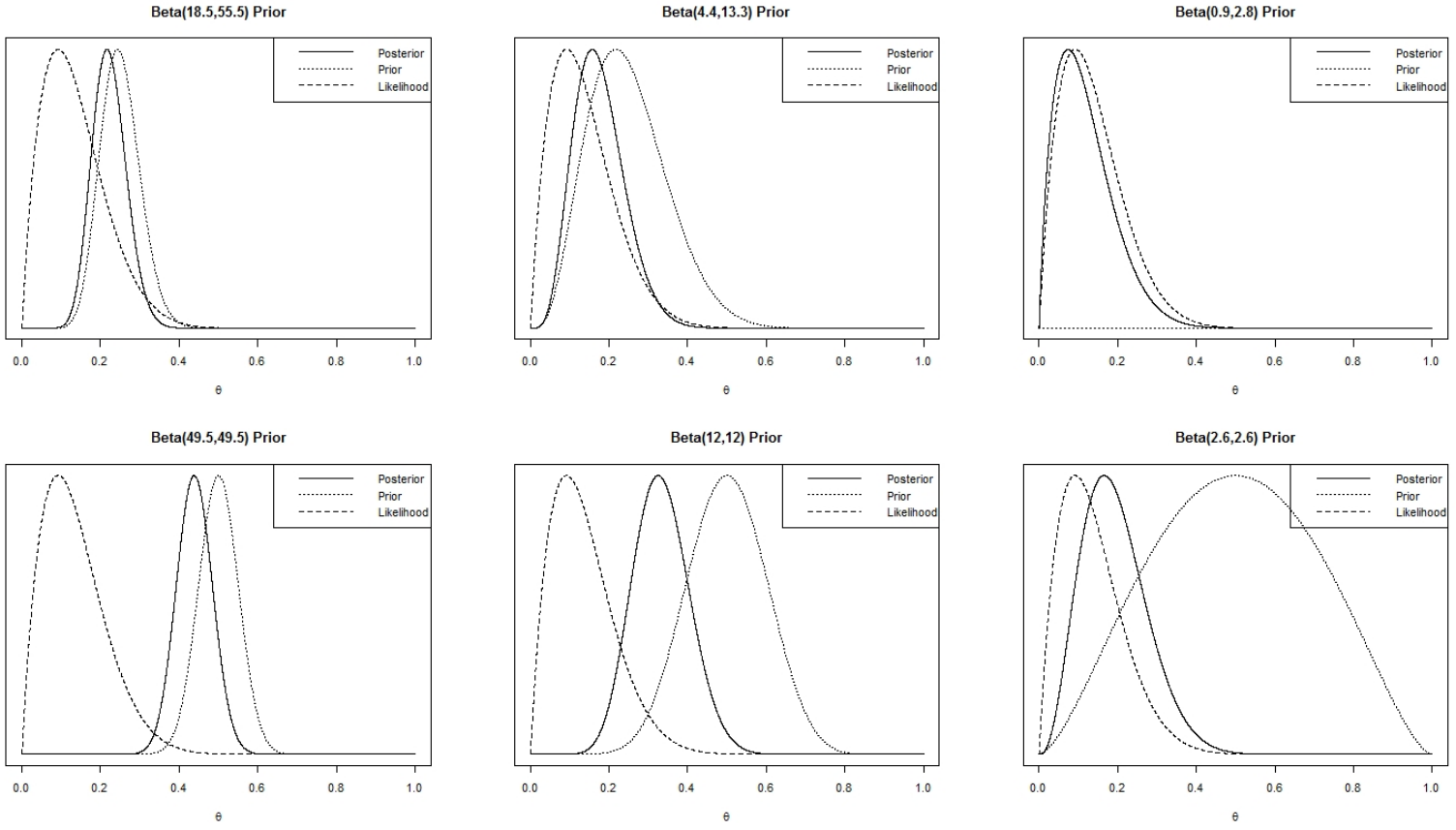

Figure 2.8: Likelihood, Prior and Posterior Distributions for Binomial Examples for $n = 20$

credible intervals are provided in Table 2.5. Given the larger Bernoulli sample size, these are considerably narrower than those in Table 2.3.

Table 2.5: Binomial Credible Intervals for $x \in [1, 3]$ and $n = 20$

Prior Distribution	*95% Credible Interval*
$beta(18.5, 55.5)$	$(0.143, 0.312)$
$beta(4.4, 13.3)$	$(0.068, 0.317)$
$beta(0.9, 2.8)$	$(0.015, 0.0.291)$
$beta(49.5, 49.5)$	$(0.351, 0.529)$
$beta(12, 12)$	$(0.2, 0.479)$
$beta(2.6, 2.6)$	$(0.058, 0.376)$

2.5 Mixed Precise and Interval Data

Suppose a daily administered oral drug has dizziness as a possible side effect within two hours of consumption. After a short period of time on the drug, a patient is asked how many days he or she experienced such dizziness. Some patients are able to provide a precise count, but others can only reply with a range of possible events. Thus, suppose after two weeks on the drug, a sample of ten patients provides the following data.

Table 2.6: Patients Reporting Drug Side Effects

Patient	*Days Dizzy*	*Patient*	*Days Dizzy*
1	2	6	3-6
2	3	7	3
3	3-4	8	4
4	3	9	5
5	2-3	10	2-4

When we have a mixture of precise and interval data as seen in Table 2.6, we have two alternatives for the likelihood. We can either incorporate both types of data into our likelihood function, or we can combine the data into one censoring interval. Here we consider both options.

Suppose we combine the data into one censoring interval. Of the 140 patient-days, dizziness was reported in at least 30 $(20 + 3 + 2 + 3 + 2)$ and as many as 37 $(20 + 4 + 3 + 6 + 4)$ days. Thus, the likelihood becomes

$$L(\theta|140, 30, 37) = \sum_{x=30}^{37} \binom{140}{x} \theta^x (1-\theta)^{140-x}.$$

Using this likelihood, we find that the MLE for θ is 0.29 and the likelihood interval is $(0.167, 0.321)$. We also calculate the Bayesian estimates using a $beta(1, 1)$ prior distribution. We find the posterior mean is 0.2429 and the credible interval is $(0.17, 0.323)$.

Alternatively, we can incorporate both precise and interval data into the likelihood function. To do this, we multiply the precise likelihood and the interval censored likelihood together:

$$L(\theta|n, \mathbf{y}, \mathbf{j}, \mathbf{k}) = \left[\prod_{i=1}^{p} \sum_{x=j_i}^{k_i} \binom{n}{x} \theta^x (1-\theta)^{n-x}\right] \left[\prod_{i=1}^{p} \binom{n}{y_i} \theta^{y_i} (1-\theta)^{n-y_i}\right].$$

Using numerical methods, we find that the MLE and likelihood interval are 0.2359 and $(0.169, 0.313)$, respectively. Using a $beta(1, 1)$ prior, the corresponding posterior mean and credible interval are 0.2398 and $(0.171, 0.316)$. We find that the point and interval estimates derived when using the interval-censored likelihood are very similar to the point and interval estimates derived when using both precise and interval data, for this example.

2.6 Other Distributions

We have derived frequentist and Bayesian interval estimates for the binomial distribution with censored counts. We now consider the censored data problem for

the negative binomial and Poisson distributions, and calculate interval estimates for multiple examples using these two distributions.

2.6.1 Negative Binomial Distribution

Let x have a negative binomial distribution with parameters θ and r, denoted $NegBin(\theta, r)$, with probability mass function

$$f(x) = \binom{x-1}{r-1}\theta^r(1-\theta)^{x-r},$$

where x is the number of trials until the r^{th} success and θ is the probability of success. The censored likelihood function is

$$L(\theta|j \leq x \leq k) = \sum_{x=j}^{k} \binom{x-1}{r-1}\theta^r(1-\theta)^{x-r}. \tag{2.18}$$

Suppose we now let $\theta \sim beta(a, b)$ so that the posterior distribution is

$$\begin{aligned}\pi\left(\theta|j \leq x \leq k\right) &\propto \left[\sum_{x=j}^{k} \binom{x-1}{r-1}\theta^r(1-\theta)^{x-r}\right]\left[\frac{1}{B(a,b)}\theta^{a-1}\left(1-\theta\right)^{b-1}\right] \\ &= \sum_{x=j}^{k} \frac{1}{B(a,b)}\binom{x-1}{r-1}\theta^{r+a-1}(1-\theta)^{x-r+b-1}. \end{aligned} \tag{2.19}$$

The marginal distribution is

$$\begin{aligned} m(j,k) &= \int_0^1 \sum_{x=j}^{k} \frac{1}{B(a,b)}\binom{x-1}{r-1}\theta^{r+a-1}(1-\theta)^{x-r+b-1}d\theta \\ &= \frac{1}{B(a,b)}\sum_{x=j}^{k}\binom{x-1}{r-1}\int_0^1 \theta^{r+a-1}(1-\theta)^{x-r+b-1}d\theta \\ &= \frac{1}{B(a,b)}\sum_{x=j}^{k}\binom{x-1}{r-1}B(r+a, x-r+b). \end{aligned} \tag{2.20}$$

The p^{th} posterior moment is

$$\begin{aligned} E\left(\theta^p|x\right) &\propto \int_0^1 \sum_{x=j}^{k} \frac{1}{B(a,b)}\binom{x-1}{r-1}\theta^p\theta^{r+a-1}\left(1-\theta\right)^{x-r+b-1}d\theta \\ &= \sum_{x=j}^{k}\binom{x-1}{r-1}\int_0^1 \theta^{p+r+a-1}\left(1-\theta\right)^{x-r+b-1}d\theta \\ &= \sum_{x=j}^{k}\binom{x-1}{r-1}B(p+r+a, x-r+b). \end{aligned}$$

Dividing by the marginal distribution in (2.20), we obtain

$$E\left(\theta^p|x\right) = \frac{\sum_{x=j}^{k} \binom{x-1}{r-1} B(p+r+a, x-r+b)}{\sum_{x=j}^{k} B(r+a, x-r+b)} \tag{2.21}$$

As an example, suppose we observe $x \in [7, 10]$ where $r = 1$. We plot this likelihood and a cutoff value of $c = 0.15$ in Figure 2.9.

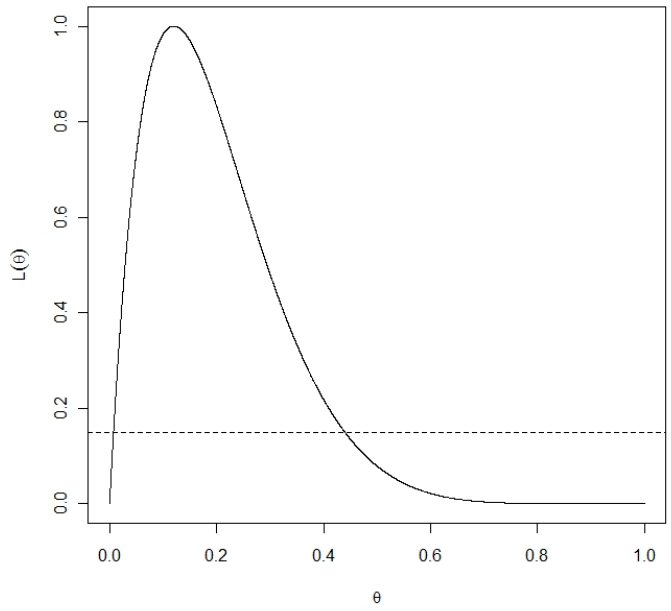

Figure 2.9: Negative Binomial Likelihood where $x \in [7, 10]$ and $r = 1$

Using numerical methods, we find that the MLE of this distribution, $\hat{\theta}$, is 0.12 and using that, find the likelihood interval to be $(0.007, 0.44)$. If we place a $beta(1, 1)$ prior on θ, and using (2.21), we calculate a posterior mean of 0.198 and a Bayesian credible interval of $(0.027, 0.485)$. Both intervals have similar coverage, but the Bayesian interval is slightly wider.

We now look at several examples of posterior means and credible intervals using the negative binomial likelihood displayed in Figure 2.9. We will consider the six beta distributions listed in Table 2.1. The likelihood function, prior distributions,

and resulting posterior distributions can be seen in Figure 2.10. The six different posterior credible intervals are listed in Table 2.7.

Table 2.7: Negative Binomial Posterior Means and Credible Intervals

Prior Distribution	*Posterior Mean*	*95% Credible Interval*
$beta(18.5, 55.5)$	0.237	$(0.027, 0.485)$
$beta(4.4, 13.3)$	0.209	$(0.078, 0.382)$
$beta(0.9, 2.8)$	0.16	$(0.020, 0.409)$
$beta(49.5, 49.5)$	0.473	$(0.379, 0.568)$
$beta(12, 12)$	0.408	$(0.245, 0.582)$
$beta(2.6, 2.6)$	0.272	$(0.078, 0.537)$

The Bayesian model performs as expected, with more shrinkage toward the MLE as the prior variability increases.

2.6.2 Poisson Distribution

Let $x \sim Poisson(\theta)$ with probability mass function

$$f(x) = \frac{\theta^x \exp(-\theta)}{x!}.$$

The censored likelihood function is

$$L\left(\theta | j \leq x \leq k\right) = \sum_{x=j}^{k} \frac{\theta^x \exp(-\theta)}{x!}. \tag{2.22}$$

Suppose we now let $\theta \sim gamma(\alpha, \beta)$ so that the posterior distribution is

$$\begin{aligned} \pi\left(\theta | j \leq x \leq k\right) &= \left[\sum_{x=j}^{k} \frac{\theta^x \exp(\theta)}{x!}\right] \left[\frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)\right] \\ &= \sum_{x=j}^{k} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\theta^{x+\alpha-1} \exp(-\theta - \beta\theta)}{x!} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \frac{\theta^{x+\alpha-1} \exp(-\theta(\beta+1))}{x!}. \end{aligned} \tag{2.23}$$

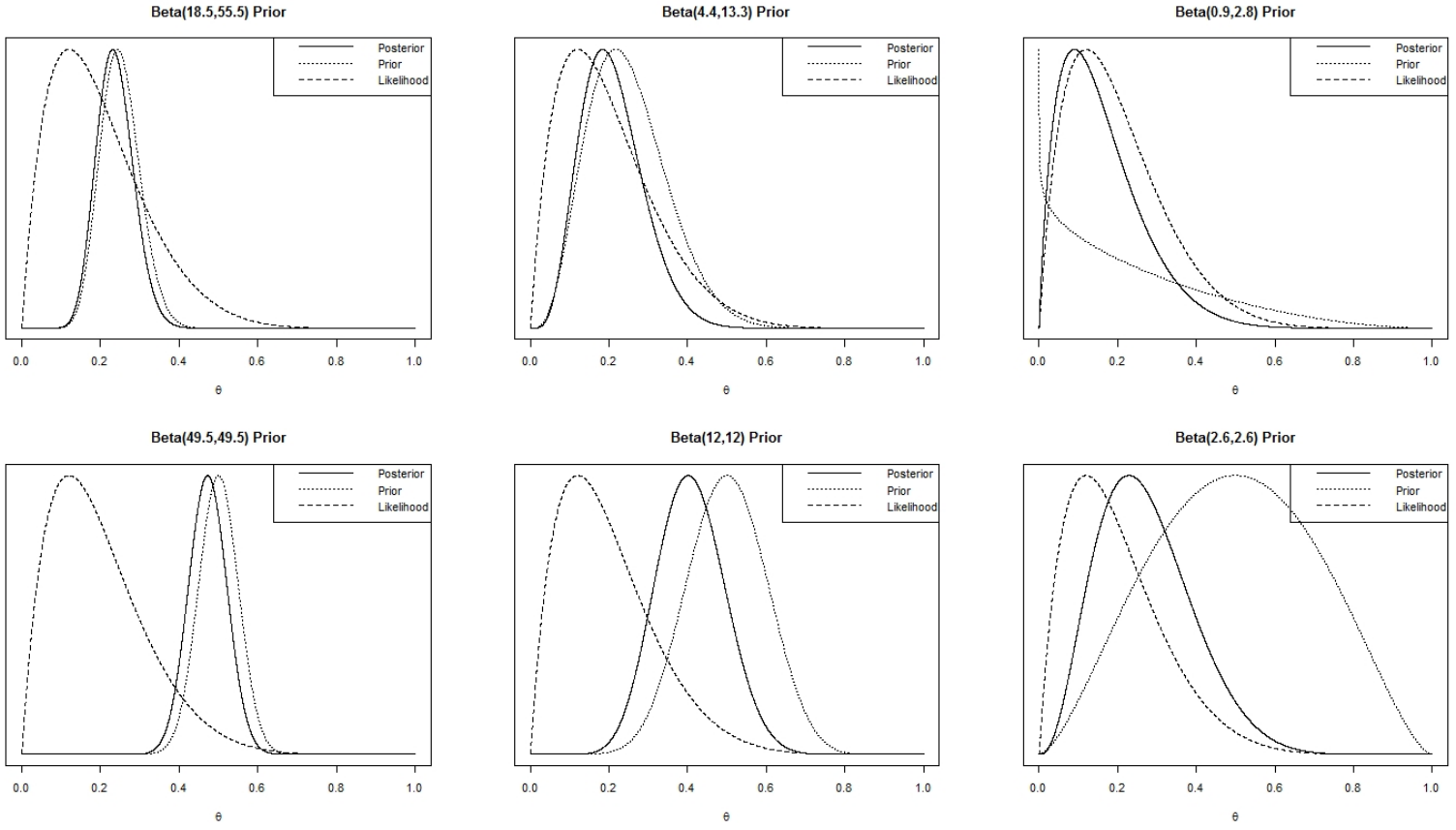

Figure 2.10: Likelihood, Prior and Posterior Distributions for Negative Binomial Examples

The marginal distribution is

$$\begin{aligned} m(j,k) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \frac{\theta^{x+\alpha-1}\exp(-\theta(\beta+1))}{x!} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \int_0^\infty \frac{\theta^{x+\alpha-1}\exp(-\theta(\beta+1))}{x!} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}} \frac{1}{x!}. \end{aligned} \tag{2.24}$$

The p^{th} posterior moment is

$$\begin{aligned} E\left(\theta^p | j \le x \le k\right) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \int_0^\infty \frac{\theta^p \theta^{x+\alpha-1}\exp\left(-\theta(\beta+1)\right)}{x!} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \frac{1}{x!} \int_0^\infty \theta^{x+\alpha+p-1}\exp\left(-\theta(\beta+1)\right) d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{x=j}^{k} \frac{1}{x!} \frac{\Gamma(x+\alpha+p)}{(\beta+1)^{x+\alpha+p}}. \end{aligned}$$

Dividing by the marginal, we obtain

$$E\left(\theta^p | j \le x \le k\right) = \frac{\sum_{x=j}^{k} \frac{1}{x!} \frac{\Gamma(x+\alpha+p)}{(\beta+1)^{x+\alpha+p}}}{\sum_{x=j}^{k} \frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}} \frac{1}{x!}}. \tag{2.25}$$

Suppose we observe the interval $[3, 5]$. The resulting likelihood and cutoff at $c = 0.15$ are displayed in Figure 2.11. Using numerical methods, we find the MLE and likelihood interval for θ to be 3.9 and $(1.1, 9.5)$, respectively.

We now calculate Bayesian credible intervals for this Poisson likelihood using four different prior distributions. Suppose we have reasonable information that the true value of θ is approximately 5. We allow the mean of the prior distribution to be 5 and construct a prior distribution whose mode is 4. We then consider the scenario where we believe the true value of θ is 2. We construct a prior distribution whose mean is 2 and mode is 1. For the last two prior distributions, we consider the possibility that the true value of θ is 10 and construct prior distributions using standard deviations 1 and 2. A summary of these prior distributions is given in Table 2.8.

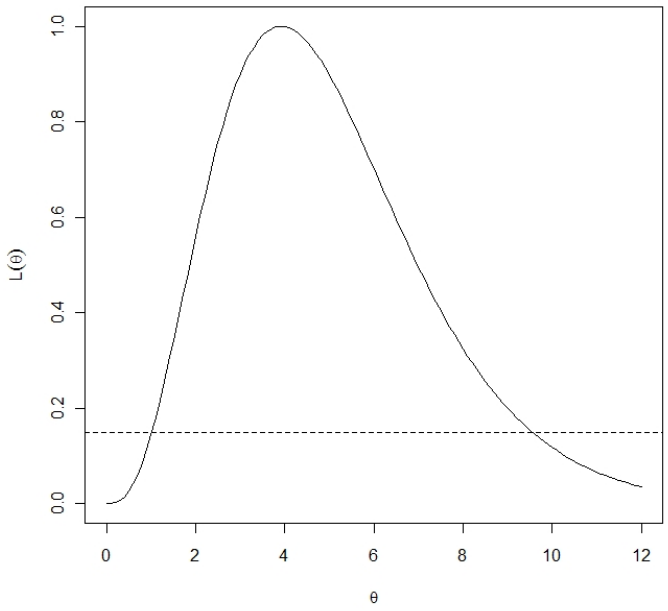

Figure 2.11: Poisson Likelihood where $x \in [3, 5]$

Table 2.8: Prior Distributions for Poisson Examples

Prior Distribution	*Prior Mean*	*Prior SD*
$gamma(2, 1)$	2	1.41
$gamma(5, 1)$	5	2.24
$gamma(100, 10)$	10	1
$gamma(25, 2.5)$	10	2

We plot each of these prior distributions, the likelihood function, and the resulting posterior distributions in Figure 2.12.

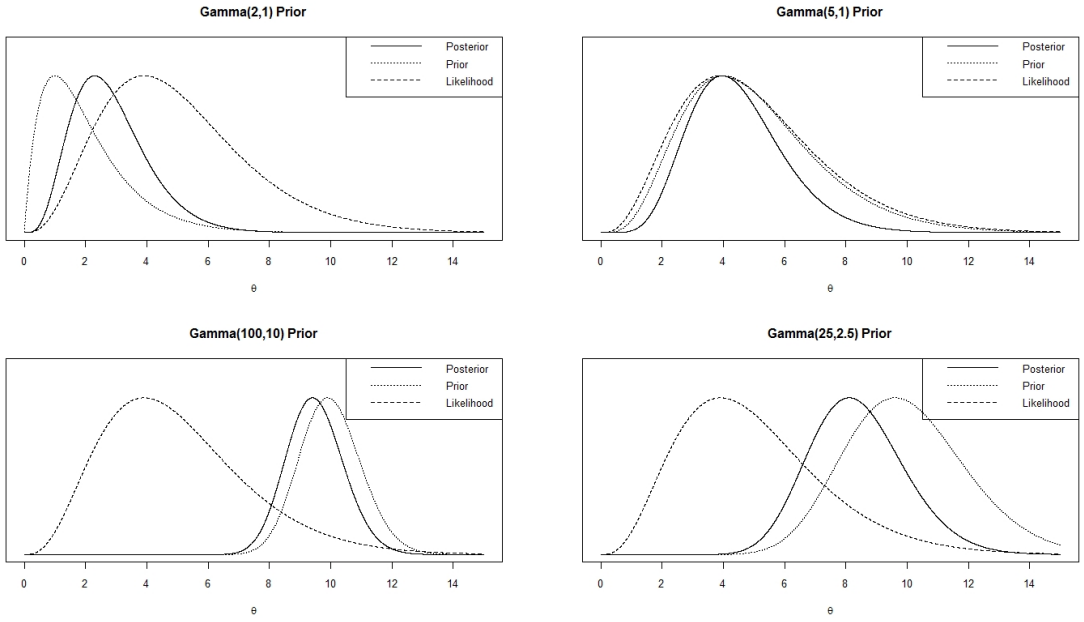

Figure 2.12: Likelihood, Prior and Posterior Distributions for Poisson Examples

The four different posterior means and credible intervals are listed in Table 2.9. As expected, posterior means shrink toward the MLE as prior standard deviations increase and posterior interval width decreases as prior standard deviations decrease.

Table 2.9: Poisson Means and Credible Intervals

Prior Distribution	*Posterior Mean*	*95% Credible Interval*
$gamma(2, 1)$	2.84	$(0.93, 5.762)$
$gamma(5, 1)$	4.48	$(1.962, 7.968)$
$gamma(100, 10)$	9.495	$(7.74, 11.39)$
$gamma(25, 2.5)$	8.39	$(5.54, 11.67)$

CHAPTER THREE

Small Sample Interval Sampling for the Binomial Distribution

In the previous chapter, we derived both frequentist and Bayesian point and interval estimates for three discrete distributions where censored sampling was an issue. We considered several individual examples using the binomial, negative binomial, and Poisson distributions. We noted several things regarding these examples. For the binomial distribution, the results obtained using the MLE derived in Frey and Marrero (2008) and the posterior mean that we derived were comparable. Also, we noted that the frequentist and Bayesian interval estimates were markedly similar although the credible intervals tended to be slightly narrower than the corresponding likelihood intervals.

In this chapter, we conduct a simulation experiment in order to compare the performance of the Bayesian and frequentist estimators based on censored binomial counts. We compare the accuracy of the MLE and posterior mean as well as the width and coverage of the likelihood and credible intervals. In our simulation, we consider seven different true values of the parameter, θ. We look at how these different values of θ and varying Bernoulli sample sizes affect the point and interval estimates.

3.1 Data Generation

The likelihood function for the censored binomial distribution is

$$L\left(\theta|n,j,k\right)=\sum_{x=j}^{k}\binom{n}{x}\theta^{x}\left(1-\theta\right)^{n-x}. \tag{3.1}$$

As a function of θ for fixed n, j, and k, this is the likelihood function of θ. However, if θ and n are both fixed, then (3.1) becomes the joint distribution of j and k.

Therefore, (j, k) has the distribution

$$f(j, k|n, \theta) = \sum_{x=j}^{k} \binom{n}{x} \theta^x (1-\theta)^{n-x} . \tag{3.2}$$

We generate values of (j, k) using a multinomial distribution with cells whose probabilities are generated by (3.2). The support of this multinomial distribution must be such that $j \leq k-1$ where $j \geq 0$ and $k \leq n$. Consider the case where $n = 5$, as depicted in Figure 3.1.

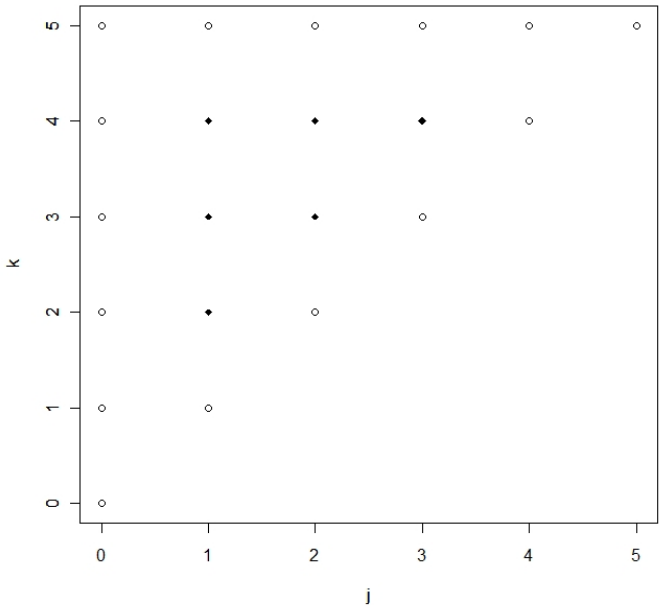

Figure 3.1: Cells of the Multinomial Distribution for $n = 5$

The upper triangular region holds all the possible combinations of j and k that could constitute the censoring interval. There are $\binom{n-1}{2}$ pairs in the support of this distribution. As we discussed in detail in the previous chapter, if $j = 0$ or $k = n$, the resulting likelihood is irregular. In order to restrict the simulations to those producing regular likelihood functions, we exclude any values of j and k that lie on the boundaries of this upper triangular region. Therefore, only pairs lying in

the interior of the triangle displayed in Figure 3.1 would be considered in a simulation for the case of $n = 5$.

We are able to compute the probability of each pair in the support of this multinomial distribution using (3.2). In order to generate values of (j, k) that will produce regular likelihood functions, we need a distribution for

$$(j, k)|0 < j \leq k - 1 < n.$$

In order to produce this distribution, we must divide (3.2) by the marginal probability

$$\begin{aligned} m &\equiv P\left(0 < j \leq k - 1 < n|\theta, n\right) \\ &= \sum_{k=2}^{n-1} \sum_{j=1}^{k-1} \sum_{x=j}^{k} \binom{n}{x} \theta^x \left(1 - \theta\right)^{n-x}. \end{aligned}$$

In the case of $n = 5$, this marginal probability is the sum of the probabilities in the interior of the upper triangular region displayed in Figure 3.1.

Suppose $n = 5$ and that we set $\theta = 0.25$. For this example, there are six possible combinations of j and k that would produce a regular likelihood function. These pairs and their conditional probabilities are displayed in Table 3.1.

Table 3.1: Multinomial distribution of (j, k) for $n = 5$ and $\theta = 0.25$

j	k	$f(j, k\|n = 5, \theta = 0.25)/m$	*Simulated Proportion of* (j, k)
1	2	0.22	0.2194
1	3	0.25	0.2482
1	4	0.26	0.257
2	3	0.12	0.1175
2	4	0.12	0.1262
3	4	0.03	0.0317

Table 3.1 also reports the proportion of pairs in each category for a simulated sample of size 10,000. The simulation frequencies are very close to what the theoretical probabilities would predict for 10,000 simulated pairs.

Before we begin the simulation experiment, we consider the plots of the six possible likelihood functions that could be produced in the simulation. According to the multinomial distribution in Table 3.1, the intervals $[1, 3]$ and $[1, 4]$ have the greatest probability of being selected in the simulation when the true value of the parameter, θ is 0.25. However, when we look at the likelihood plots in Figure 3.2, the MLE is approximately 0.4 when $X \in [1, 3]$ and 0.5 when $X \in [1, 4]$. Therefore, we should expect to see some degree of bias in our simulated point estimates, especially when using a uniform prior on θ.

3.2 Simulation Methodology

In our simulation, we calculate the maximum likelihood estimate, posterior mean, posterior standard deviation, likelihood interval, and credible interval for various values of θ, Bernoulli sample sizes, and censoring interval widths. The calculations for the MLE, posterior mean, and posterior standard deviation use the development from Chapter 2, specifically equations (2.3), (2.14), and (2.15).

While simulating the point estimates is fairly straightforward, simulating the interval estimates is more complicated. We first consider the simulation of the likelihood interval. We create a vector of 10,000 equally spaced values of θ ranging between 0 and 1. For each of these values of θ, we calculate the likelihood:

$$L\left(\theta | j \leq X \leq k\right) = \sum_{x=j}^{k} \binom{n}{x} \theta^{x} \left(1-\theta\right)^{n-x}.$$

This calculation results in 10,000 points of the likelihood function. Recall that the likelihood interval consists of all values of θ that satisfy

$$\frac{L\left(\theta | j \leq x \leq k\right)}{L\left(\hat{\theta} | j \leq x \leq k\right)} > c.$$

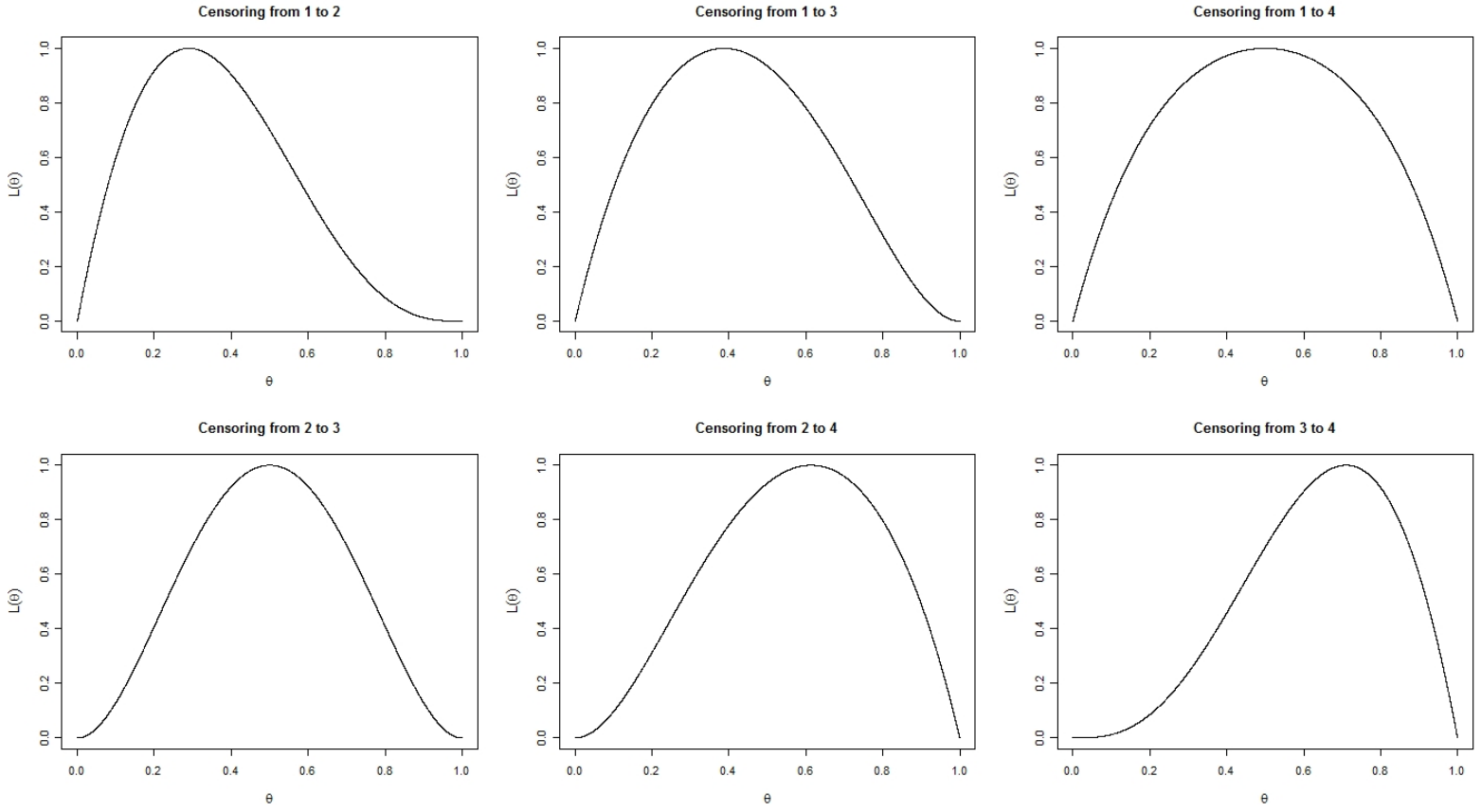

Figure 3.2: Likelihood Functions for $n = 5$

Since we have assured ourselves of obtaining a regular likelihood in the simulation, we can obtain a 95% interval by setting $c = 0.15$. Therefore, we collect all values of θ that are associated with values of the normalized likelihood that are greater than 0.15 and place them into a vector. The minimum and maximum value in this vector are the lower and upper bound of the likelihood interval, respectively.

We now consider the calculation of the Bayesian credible interval. As discussed in the previous chapter, the limits of the 95% credible interval are found by solving

$$\int_0^y \pi\left(\theta|n,j,k\right) = 0.025 \tag{3.3}$$

and

$$\int_0^z \pi\left(\theta|n,j,k\right) = 0.975. \tag{3.4}$$

In the simulation, we place a $beta(1,1)$ prior distribution on θ. Therefore, the posterior distribution will be

$$\pi\left(\theta|n,j,k\right) = \frac{\sum_{x=j}^{k} \binom{n}{x} \theta^x \left(1-\theta\right)^{n-x}}{\sum_{x=j}^{k} \binom{n}{x} B(x+1, n-x+1)}.$$

Note that this distribution is a polynomial in θ. Consequently, the integrations needed to construct interval estimates are straight forward. For example, consider the posterior for $n = 5$ and the observation $[1, 2]$. The resulting posterior is

$$\pi\left(\theta|x\right) = \frac{\binom{5}{1}\theta^1(1-\theta)^4 + \binom{5}{2}\theta^2(1-\theta)^3}{\sum_{x=1}^{2} \binom{n}{x} B(x+1, 5-x+1)} \tag{3.5}$$

$$= h[\theta - 2\theta^2 + 2\theta^4 - \theta^5], \tag{3.6}$$

where

$$h = \frac{15}{\sum_{x=1}^{2} \binom{n}{x} B(x+1, n-x+1)}.$$

Because the posterior can be written as a polynomial, we are able to obtain the exact integral of the posterior. For example,

$$\int_0^y \pi(\theta|n,j,k)d\theta = h[\theta - 2\theta^2 + 2\theta^4 - \theta^5].$$

Therefore, the solution (3.3) and (3.4) is the solution to

$$h[\theta - 2\theta^2 + 2\theta^4 - \theta^5] = 0.025$$

and

$$h[\theta - 2\theta^2 + 2\theta^4 - \theta^5] = 0.975.$$

By using this method, finding the limits of the credible interval becomes a matter of finding roots of polynomials.

3.3 Simulation Results

In the simulation, we consider the effect of sample size and the true value of the parameter, θ, on the point and interval estimates. Bernoulli sample sizes of 5, 10, and 20 are studied in the simulation. We also consider seven different true values of θ: 0.05, 0.10, 0.25, 0.5, 0.75, 0.90, and 0.95. For the Bayesian estimates, we place a $beta(1,1)$ prior distribution on θ. In addition, we investigate the extent to which restricting the width of the censoring interval to 3 improves point and interval estimates. Each simulation consists of 10,000 replications. We then calculate the mean and standard deviation of those 10,000 estimates.

We first consider the case where $\theta = 0.05$ and the censoring interval width is unrestricted. The simulation results are presented in Figure 3.3 and in Table A.1. In this figure, the horizontal bars represent the means of the lower and upper bounds of the likelihood and credible intervals produced in the simulation. The points represent the means of the MLEs and the posterior means. The variability in these estimates is shown by the grey boxes, which cover plus or minus one simulation standard deviation.

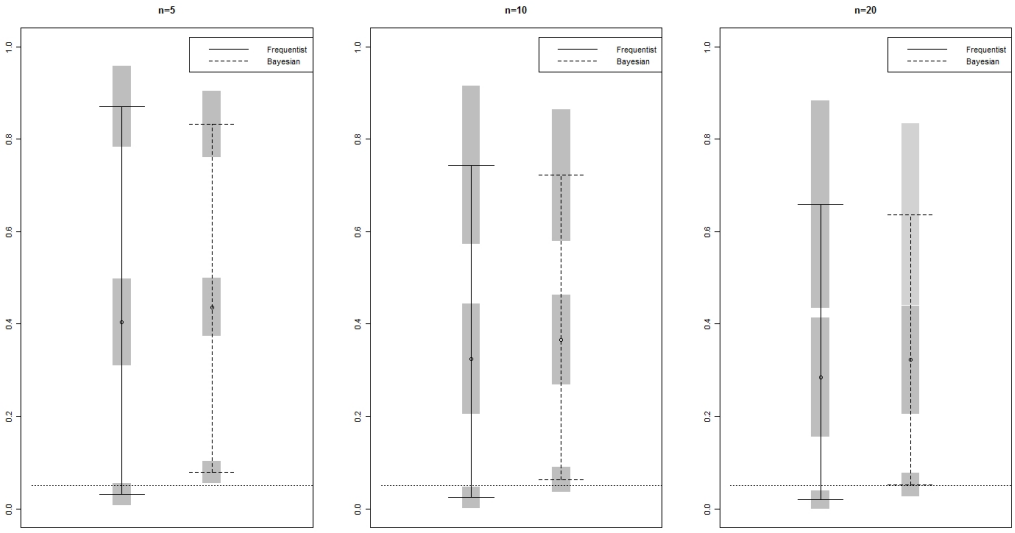

Figure 3.3: Simulation Results for $\theta = 0.05$ - Unrestricted Censoring Intervals

As expected, both the MLE and the posterior mean are extremely biased. As the sample size increases, the point estimates do begin to decrease, but they do not come anywhere close to the true value of the parameter. Estimation of small probabilities with small samples is, of course, problematic in general and not helped by the presence of interval censoring. We also considered the case of $\theta = 0.01$ and found similar results.

Interval estimation of θ, both with frequentist and Bayesian methods, is also poor in this small probability case. Both types of intervals are very wide and improve only marginally as n increases.

In addition to the problems with the bias, note that the simulation standard deviation increases as the Bernoulli sample size increases. In order to investigate this, we consider the distribution of the censoring interval width, $k - j$, for all three Bernoulli sample sizes. We also calculate the mean and standard deviation of $k - j$.

The relative frequency distribution for 10,000 replications of $k - j$ when $n = 5$ are displayed below.

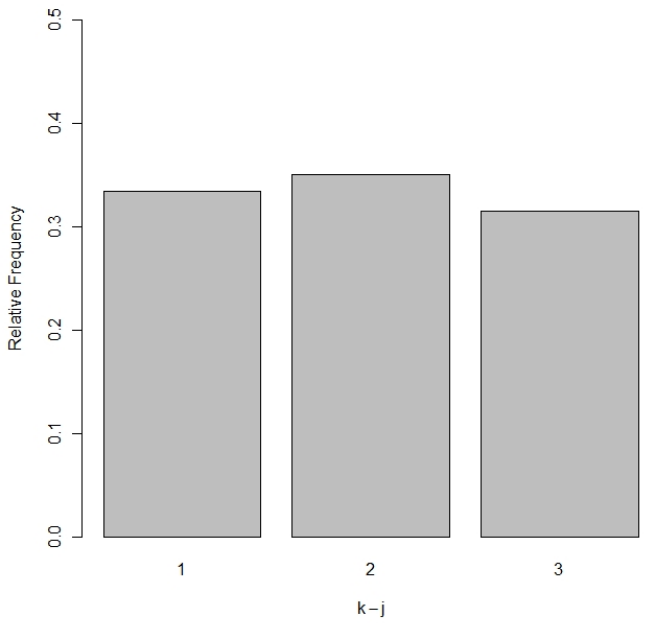

Figure 3.4: Relative Frequency Distribution of $k - j$ when $n = 5$

The mean and standard deviation for $k - j$ when $n = 5$ are 1.9804 and 0.805, respectively. Based on the results in Figure 3.4, the the distribution of $k - j$ appears to be nearly uniform. We consider the distribution of $k - j$ for $n = 10$.

The mean and standard deviation of $k - j$ when $n = 10$ are 4.366 and 2.233, respectively. The distribution of $k - j$ appears to be nearly uniform. We next look at the distribution of $k - j$ when $n = 20$. The results are below.

The mean and standard deviation of $k - j$ when $n = 20$ are 9.267 and 5.065, respectively. Again, the distribution of $k - j$ is roughly uniform. This distribution explains the increasing simulation standard deviation. As n increases, the support

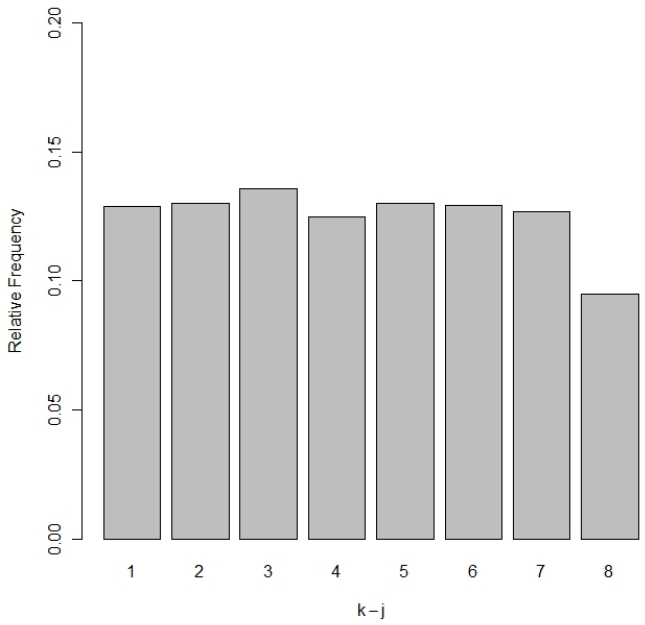

Figure 3.5: Relative Frequency Distribution of $k - j$ when $n = 10$

of the distribution of (j, k) in (3.2) increases in cardinality and with nearly equal probability for the added pairs.

Because of this association between simulation standard deviation and the censoring interval width, we consider restricting the width of the censoring interval. We select 3 as the maximum width of the interval because 3 is the largest width available for the case where $n = 5$.

We now consider the case where $\theta = 0.05$ and the censoring interval is restricted to 3. The results for this simulation are seen in Figure 3.7 and Table A.2. There are several notable differences in the simulation results when the censoring interval width is restricted to 3. First, all of the likelihood and credible intervals contain $\theta = 0.05$. In addition, all of the intervals are much narrower than they were when the censoring interval width was unrestricted. This is particularly true for $n = 20$. However, it is interesting to note that the interval width of the Bayesian and likelihood intervals does not seem to differ significantly.

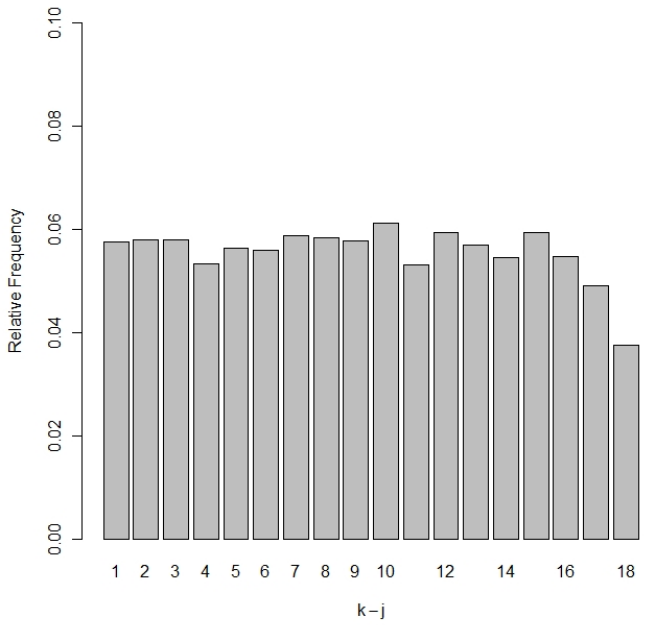

Figure 3.6: Relative Frequency Distribution of $k - j$ when $n = 20$

The point estimates are still biased; however, at $n = 20$, the MLE and the posterior mean are 0.108, and 0.148, respectively. The bias appears to have decreased significantly when we restrict the censoring interval. By restricting the censoring interval to 3, we are eliminating the majority of the possible censoring intervals when $n = 10$ and $n = 20$. When we do this, we are eliminating intervals that contain values of x unlikely to produce $\theta = 0.05$. Therefore, narrower intervals and less biased point estimates are to be expected.

We now consider the case for $\theta = 0.25$ and the censoring interval width is unrestricted. The simulation results are seen in Figure 3.8 and Table A.5.

Here all of the intervals contain the true value of the parameter, $\theta = 0.25$. Furthermore, all of the Bayesian credible intervals are slightly narrower than the likelihood intervals, and both intervals become progressively narrower as the Bernoulli sample size increases. Both the MLE and posterior mean are still positively biased, but again, that is not surprising.

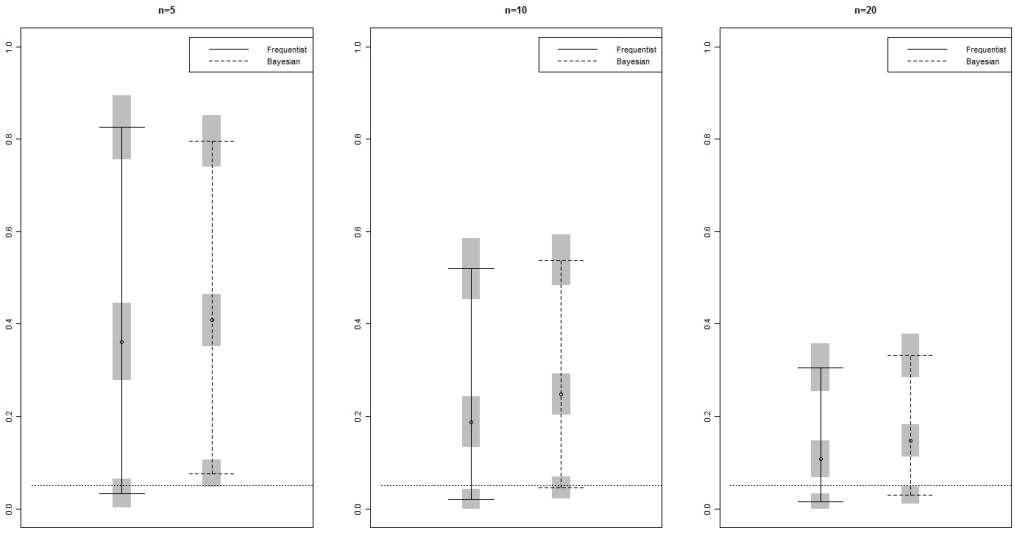

Figure 3.7: Simulation Results for $\theta = 0.05$ - Restricted Censoring Intervals

We now consider the case of $\theta = 0.25$ where the width of the censoring interval is restricted to 3. The simulation results are displayed in Figure 3.9 and Table A.6.

The resulting likelihood and credible intervals contain the true value of θ. Not surprisingly, the intervals produced from the restricted censoring intervals are significantly narrower than the intervals produced from the unrestricted censoring intervals. The likelihood and credible intervals did not appear to differ significantly in width and coverage, again not surprising given that we have used a $beta(1, 1)$ prior. In addition, the bias in the point estimates decreases as the sample size increases. When $n = 20$, the MLE and posterior mean are very close to 0.25.

There is clearly a problem with bias in the point estimates. We consider the case where $\theta = 0.25$. Suppose we think, *a priori*, that θ is less than 0.40 with high probability. We construct a prior distribution, both with a mean of 0.3 and a standard deviation of 0.21. Using this criteria, we obtain the following prior distribution: $beta(1.18, 2.76)$.

We calculate posterior means using this $beta(1.19, 2.76)$ prior distributions. We consider the case where $n = 20$ and we have the observation $[1, 4]$. We calculate

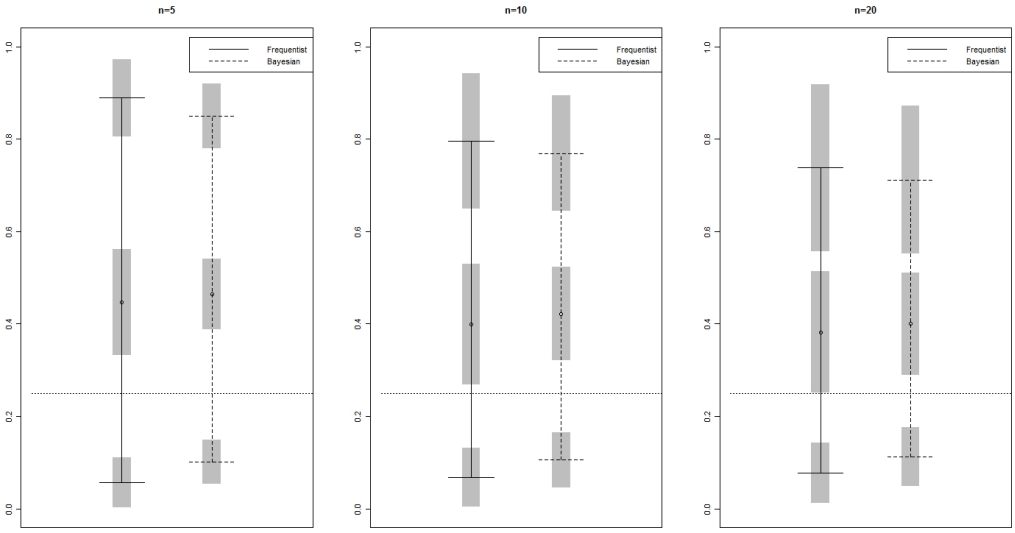

Figure 3.8: Simulation Results for $\theta = 0.25$ - Unrestricted Censoring Intervals

a posterior mean of 0.152 when we use the $beta(1.19, 2.76)$ prior distribution. As this example illustrates, it is possible to reduce the large bias present in the point estimates by using somewhat more informative prior distributions.

Although the point estimates are biased in these simulations, the majority of the intervals do contain the true value of the parameter. In addition, the intervals become narrower and the point estimates becomes slightly less biased as the Bernoulli sample size is increased. This can also be seen in the simulation results for $\theta = 0.10$, 0.50, 0.75, and 0.90, and 0.95. These results can be found in Appendix A.

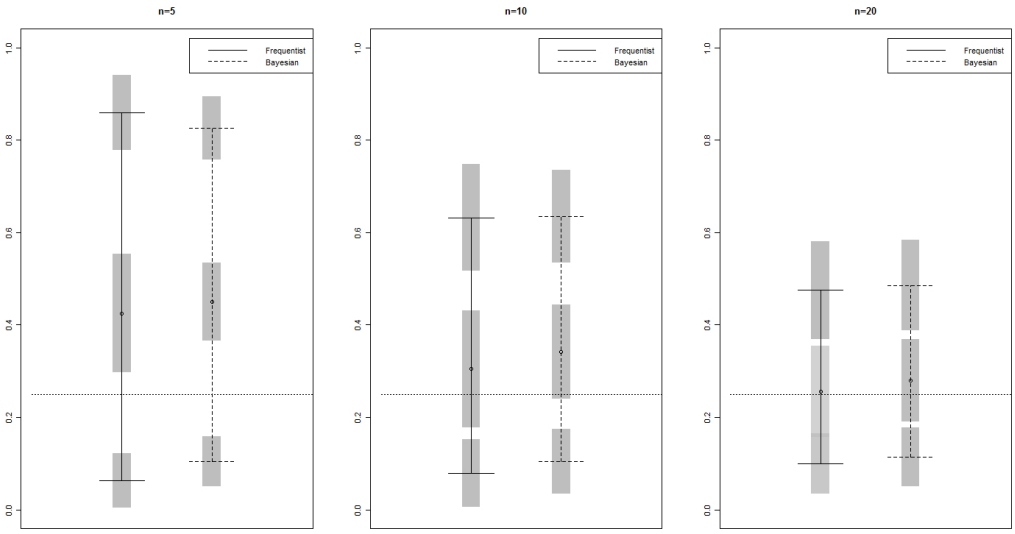

Figure 3.9: Simulation Results for $\theta = 0.25$ - Restricted Censoring Intervals

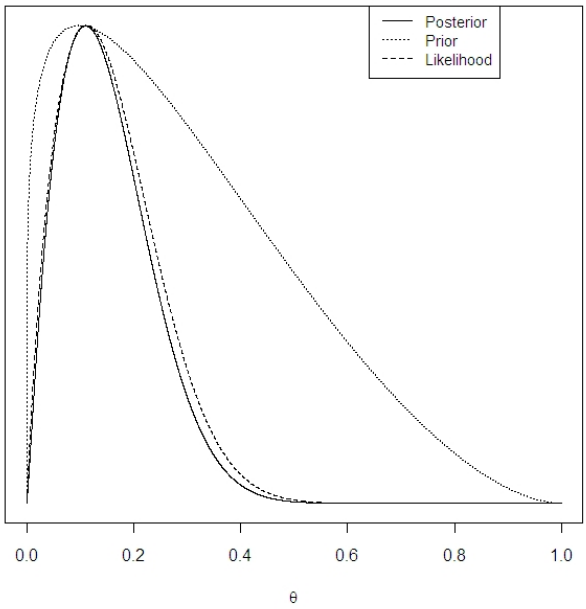

Figure 3.10: Posterior Distributions for $n = 20$ and $x \in [1, 4]$

CHAPTER FOUR

The Effect of Bioavailability on the Continual Reassessment Method

The primary goal of a Phase I trial is to accurately determine the appropriate dose of a compound for use in subsequent phases of the clinical trial. Here the focus is on safety rather than efficacy. A safe dosage level will be estimated by increasing this dose from a nominal, safe level that can be tolerated by members of a sample from the target population. This dose is commonly known in the literature as the maximum tolerated dose (MTD). It is assumed that the compound's efficacy will increase with observed toxicity.

Several statistical methods have been designed for the purpose of estimating the MTD. Among the more commonly recognized methods are the standard $3 + 3$ design and the Bayesian continual reassessment method.

The more traditional method used in Phase I trials is the $3 + 3$ design. In this method, toxicity is defined as a binary event. When the $3 + 3$ design is utilized, patients are treated in groups of 3. The algorithm will escalate and de-escalate the allotted dose, depending on how many toxicities are experienced in each group of patients. The MTD is chosen as the highest dose assigned with the lowest toxicity rate. Due to its simplicity, this method is generally preferred by clinicians. However, it has been demonstrated that the $3 + 3$ method tends to select doses with toxicities significantly less than the desired toxicity rate. For more information on the $3 + 3$ method see, for example, Garrett-Mayer (2006).

4.1 Introduction to the Continual Reassessment Method

In order to use the continual reassessment method (CRM) for a Phase I study, the clinician must specify, prior to the start of the trial, how patients are expected

to respond to the compound. This is done either by specifying a particular level of toxicity response for each of the discrete dose levels or by specifying certain percentiles of toxicity response if there are continuous dosages available. Using this prior model and the desired rate of toxicity, we can estimate the initial dose to be administered to the first cohort of patients. Once this dose is administered to the first group of patients, their toxicity response is recorded and the model is updated using their response. This cycle will continue until either the maximum sample size is attained or a stopping rule for the trial is utilized.

The CRM utilizes statistical models that explain the relationship between dose level and experienced toxicity. Among the models that have been used to explain this relationship are the one and two parameter logistic, the one parameter power, the hyperbolic, and the arctangent models. Althugh all of these models are commonly used in the literature, the logistic models are by far the most frequently used in practice. For a more comprehensive overview of the CRM, including the advantages and disadvantages of this method, refer to Eisenhauer et al. (2006).

In this chapter, we examine the effect of bioavailability on the performance of the CRM. In Section 4.2, we introduce the concept of bioavailability and its impact on clinical trials as seen in the medical literature. In Section 4.3, we compare the performance of the one parameter logistic model of the CRM to the standard design. In Section 4.4, we discuss the power model and compare its performance to that of the standard design and the one parameter logistic model. In Section 4.5, we discuss the possibility of a maximum absorbable dose and how the knowledge of this dose could impact the performance of the CRM.

4.2 Bioavailability and the CRM

Use of the CRM requires a few assumptions regarding the relationship between the administered dose and the toxicity response. It is widely assumed in the medical

and statistical literature that efficacy will increase as the dosage increases. Since bioavailability is not accounted for in the CRM, it is also implicitly assumed that the targeted tissue receives the administered dose in full. If the drug under trial is administered intravenously, then this level of drug absorption, or bioavailability, will be 100%. However if the compound is administered orally, bioavailability will vary more from patient to patient than if the compound were administered intravenously. For more information on bioavailability, specifically in cancer studies, refer to Tozer and Rowland (2006).

It has been shown that ignoring bioavailability can severely compromise the performance of the CRM. To illustrate the effect of ignoring bioavailability, consider the administration of midazolam for the purpose of sedation before cardiac catheterization as described in Fabre et al. (1998). In this study, the CRM was used to determine the appropriate dose of midazolam to administer to infants prior to the procedure. The reported bioavailability ranged from 15% to 27%. As a result of the low bioavailability, the patients did not display the level of sedation required for the procedure, and the CRM continuously increased the dosages until the maximum available dose was recommended for 15 of the 16 patients in the trial. However, even this maximum dose was shown to be ineffective, primarily due to the low bioavailability displayed in the patients. Clearly, the CRM is not robust to the effects of bioavailability.

Oncology patients enrolled in a Phase I clinical trial are generally terminally ill, and have exhausted all of their treatment options. As a result, their exhibited bioavailability is often significantly lower than healthy patients. Consider, for example, etoposide, which is indicated for small-cell lung cancer, testicular cancer, and various lymphomas. Hande et al. (1993) studied the bioavailability of two doses of etoposide. When a dose of 100 mg was administered to 11 patients, a mean bioavailability of 76% with a standard deviation of 22% was observed. However, when a

dose of 400 mg was administered to 6 patients, a mean bioavailability of 48% with a standard deviation of 18% was observed. It is important to note in this example that bioavailability does not necessarily increase with the dosage level. In particular, dose proportionality need not hold (i.e., doubling the dose need not double the availability).

4.3 Comparing the One Parameter Logistic Model to the Standard Design

4.3.1 The CRM and the Standard Design

There has been substantial work in comparing the performance of the CRM and the $3 + 3$ design. While the $3 + 3$ design is more familiar to clinicians and much easier to understand and implement, there are distinct disadvantages to its use. There have been numerous studies indicating that it yields poor estimates of the true MTD, resulting in an incorrect dose level for the later phases of the clinical trial. In fact, it has been shown that the $3 + 3$ method will generally treat a high percentage of patients outside of the therapeutic range (Potter, 2006). As a result of this, the $3 + 3$ method often recommends low and ineffective doses for future trials. Also, the $3 + 3$ design is unable to use all the available information in its dose selection procedure. It only considers data collected from the current cohort, and ignores any responses that have been collected earlier in the trial.

In contrast, the CRM has many advantages that recommend its use for Phase I studies. It has been shown to treat a higher proportion of patients at doses closer to the correct MTD (Iasonos et al., 2008). Unlike the $3 + 3$ design, it provides an estimate of the MTD using formal statistical methods, and thus allows for the description of the uncertainty about the dose level selection. In addition to incorporating the data for the current cohort of patients and all previous cohorts, the CRM can also use any prior information and beliefs the clinicians might have about the compound under study (Garrett-Mayer, 2006).

When these methods are implemented, it is generally assumed that there is 100% bioavailability in the patient population. However, as noted, this is known to be problematic, for example, when considering oncology patients enrolled in a Phase I trial. In our first simulation, we will compare the performance of the $3+3$ method and the CRM when the effect of bioavailability is recognized.

4.3.2 The One-Parameter Logistic Model

We begin our simulation studies by comparing the one-parameter logistic model of the CRM to the standard $3+3$ trial design. The one-parameter logistic model is the most commonly used of the various CRM models in both the literature and in practice. It is defined as

$$P(y=1|x=dose) = \frac{\exp(\alpha_0+\beta x)}{1+\exp(\alpha_0+\beta x)}, \tag{4.1}$$

where $y=1$ indicates the occurrence of a severe toxic response and x is the dose level. This response is often referred to a dose limiting toxicity (DLT).

When this model is utilized, the value of the intercept is fixed at $\alpha_0 = 3$ and the slope parameter is allowed to vary. This constant intercept is chosen in order to obtain a vague prior such that the *a priori* Bayesian 95% credible intervals for the probability of dose limiting toxicity at each individual dose covers as much of the $(0,1)$ interval as possible (O'Quigley and Chevret, 1991). As a result of this constraint, the intercept is often fixed at 3 in both the literature and in practice. As can be seen in the figure below, when the intercept is fixed at 3, the 95% credible intervals do extensively cover the $(0,1)$ interval when an exponential prior on β with parameter 1 is utilized.

4.3.3 Problems with the Two-Parameter Logistic Model

One alternative to this one-parameter logistic model is the two-parameter logistic model in which both the slope and intercept parameters are allowed to vary.

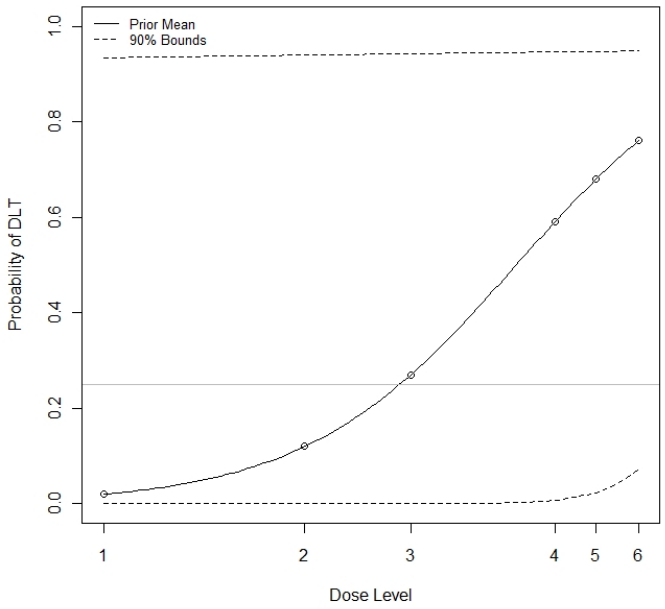

Figure 4.1: Prior Credible Intervals for the Probability of Dose Limiting Toxicity

However, this model is infrequently used for various reasons. Although the one-parameter logistic model is less flexible than the two-parameter logistic model, it often performs better in regard to the correct selection of the MTD. For example, O'Quigley et al. (1990), showed that the one-parameter model selected the correct MTD 57% of the time, and the two-parameter model selected the correct MTD only 48% of the time. Intuitively, one would think that the two-parameter model would be more accurate due its greater flexibility. However, because of the small sample sizes in a Phase I trial, there is not enough data to adequately update two parameters. Adding the extra parameter makes the model unidentifiable.

Shu and O'Quigley (2008) discuss the convergence problems associated with the two-parameter CRM. The initial cohort of patients in a Phase I clinical trial normally consists of three patients. Convergence to a single dose level is problematic when using the two-parameter model since three patients would not provide enough data to update the values for both parameters. As a result, both O'Quigley

et al. (1990) and Shu and O'Quigley (2008) strongly recommend the use of the one-parameter CRM model over the two-parameter CRM model. In addition, they recommend that additional rules be incorporated into the CRM to control for skipping dose levels. Because the CRM tends to rapidly converge to the true MTD, there is a tendency to skip multiple dose levels in that process. This is generally not allowed in practice because most clinicians would feel extremely uncomfortable skipping several dose levels of an untried compound. For our study, we will only consider the one-parameter models where dose levels can only increase in increments of one.

4.3.4 Simulating Bioavailability

In order to account for bioavailability in the CRM, we create a multiplicative factor in the one parameter dose model so that

$$P\left(DLT|x\right) = \frac{\exp\left(3+\beta\gamma x\right)}{1+\exp\left(3+\beta\gamma x\right)},$$

where β is the real-valued slope of the regression model, $x > 0$ is the administered dose, and γ is the bioavailability coefficient, which is restricted to $(0, 1]$. This coefficient, γ, represents the proportion of the drug that is absorbed into the bloodstream of the patient. The product of this coefficient and the administered dose will represent the effective dose for that patient. Using this effective dose and the true dose response model, we are able to determine the true toxicity rate of the effective dose.

In the simulation, we assume bioavailability varies among patients. To model this, we take γ to have a beta distribution with mode at the availability percentage, u. We denote this random variable by γ_u. For our simulation, we are interested in the response of the CRM to five different levels of bioavailability, u: 35%, 65%, 75%, 90%, and 99%. For each of these values of bioavailability, we construct a beta distribution with mode γ_u, from which values are drawn to represent the bioavailability of each patient in the trial. For example, consider the case of 35% bioavailability. For

this case, we constructed a distribution where the mode occurs at 0.35 and where 95% of the possible values will be greater than 0.2. Using these qualifications, we find the following distribution for $\gamma_{0.35}$:

$$\gamma_{0.35} \sim beta\,(7.3057, 12.7106)\,.$$

In the same way, we construct distributions for the other four levels of bioavailability, as summarized in Table 4.1.

Table 4.1: Beta(r,s) Distributions for Bioavailability Coefficients

Bioavailability Level	r	s
35%	7.3057	12.7106
65%	20.9967	11.7675
75%	9.6284	3.8761
90%	5.3842	1.4871
99%	88.28	1.8816

When the standard design is used in simulation, we have a set of dose levels and corresponding true probabilities of toxicity. The only way to simulate the effect of bioavailability on the standard design is to take a randomly generated value from one of the beta distributions in Table 4.1 and multiply it by the true probability of toxicity.

4.3.5 Adjusting for Bioavailability

We begin our simulation by specifying the true dose response scenario if bioavailability is 100%.

Table 4.2: True Dose Response Scenario for 100% Bioavailability

Dose Level	*Administered Dose*	*P(DLT)*
1	100	0.01
2	200	0.10
3	300	0.25
4	400	0.45
5	500	0.65
6	600	0.75

The probability of an individual patient experiencing DLT decreases as the level of bioavailability decreases. We are interested in determining what the true value of the MTD is with this change in toxic response. Using the one-parameter logistic model and the beta distributions for bioavailability listed in Table 4.1, we first consider a plot of the true dose response curve adjusted for bioavailability levels of 99%, 90%, 75%, 65%, and 35%. With the target toxicity rate set at 25%, we can see that the value of the MTD increases rapidly as the level of bioavailability decreases. When the level of bioavailability is 99%, the true value of the MTD is approximately 300 mg, or dose level 3. However, when the level of bioavailability drops to 65%, the true value of the MTD increases to approximately 500 mg, or dose level 5.

Since bioavailability will vary across subjects, the MTD will vary as well. Indeed, we can write x_{MTD} as a function of γ_u; for a target toxicity rate, p, the solution of

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta\gamma_u x_{MTD}$$

is

$$x_{MTD} = \frac{1}{\beta\gamma_u}\left[\log\left(\frac{p}{1-p}\right) - \alpha\right], \tag{4.2}$$

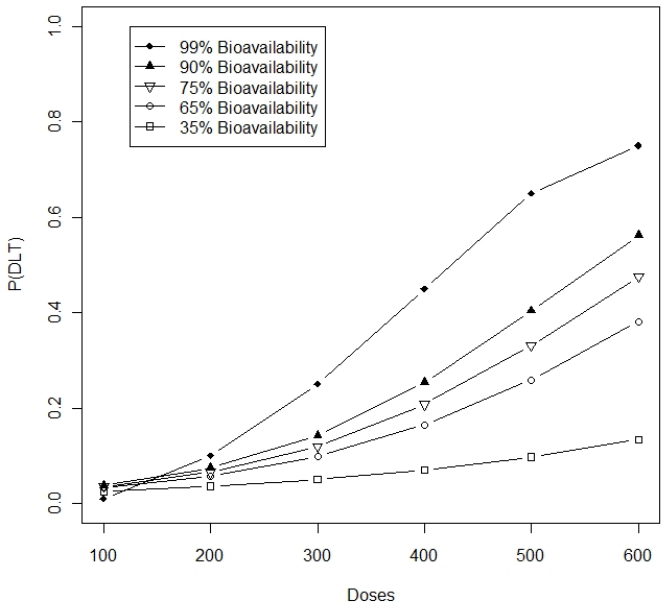

Figure 4.2: Logistic Dose Response Curves Adjusted for Bioavailability

where α and β are intercept and slope of the true dose response model, respectively. So, for fixed α, β, and p, we have

$$E\left(x_{MTD}\right) = E\left(\frac{1}{\gamma_u}\right)\frac{1}{\beta}\left[\log\left(\frac{p}{1-p}\right) - \alpha\right] \tag{4.3}$$

and

$$Var\left(x_{MTD}\right) = Var\left(\frac{1}{\gamma_u}\right)\frac{1}{\beta^2}\left[\log\left(\frac{p}{1-p}\right) - \alpha\right]^2. \tag{4.4}$$

Now, we can use the fact that (Gupta and Nadarajah, 2004) if $y \sim beta(a, b)$, then

$$E\left(\frac{1}{y^r}\right) = \frac{B(a-r,b)}{B(a,b)}, \text{ for } r = 1, 2, \ldots \text{ and } a > r. \tag{4.5}$$

Therefore,

$$\begin{aligned}\mu_{MTD} &\equiv E\left(x_{MTD}\right) \\ &= \frac{B(a-1,b)}{B(a,b)}\frac{1}{\beta}\left[\log\left(\frac{p}{1-p}\right) - \alpha\right]\end{aligned}$$

and

$$\begin{aligned}\sigma_{MTD} &\equiv \sqrt{Var\left(x_{MTD}\right)} \\ &= \left[\frac{B(a-2,b)}{B(a,b)} - \left(\frac{B(a-1,b)}{B(a,b)}\right)^2\right] \frac{1}{\beta^2}\left[\log\left(\frac{p}{1-p}\right) - \alpha\right]^2.\end{aligned}$$

Consider the case of 65% bioavailability where $\gamma_{0.65} \sim beta\left(20.99, 11.77\right)$ and $p = 0.25$. Then,

$$\begin{aligned}\mu_{MTD} &= \frac{1.589}{0.00891}\left[\log\left(\frac{0.25}{1-0.25}\right) - (-3.93841)\right] \\ &= 506.4\end{aligned}$$

and

$$\begin{aligned}\sigma_{MTD} &= \sqrt{Var\left(X_{MTD}\right)} \\ &= \sqrt{(0.049)\left(1.015 \times 10^5\right)} \\ &= 70.6.\end{aligned}$$

Using the same methodology, we calculate the expected value and standard deviation of the MTD at the other levels of bioavailability. The results are listed in Table 4.3.

Table 4.3: Expected Value of the MTD Using the Logistic Model

Level of Bioavailability	*Distribution*	*Mode*	*SD*	μ_{MTD}	σ_{MTD}
35%	beta(7.31,12.71)	0.35	0.105	960.7	341.0
65%	beta(20.99,11.77)	0.65	0.082	506.4	70.6
75%	beta(9.63,3.88)	0.75	0.118	462.0	93.4
90%	beta(5.38,1.49)	0.90	0.147	427.1	117.0
99%	beta(88.28,1.88)	0.99	0.015	332.7	12.6

4.3.6 Simulation Results

We now consider the results from our simulation. We examine the effect of five levels of bioavailability on the performance of both the one-parameter logistic model of the CRM and the standard $3 + 3$ design. Specifically, we are interested in the probability of selecting the expected MTD and the average number of patients assigned to each dose in a hypothetical Phase I trial.

For each of the five levels of bioavailability, the simulation was replicated 500 times. This represents the occurrence of 500 Phase I trials. We begin with 99% bioavailability because the results we see should be similar to those obtained in the 100% case. We then consider smaller levels of bioavailability. The results for the case of 99% bioavailability can be seen in Figures 4.3 and 4.4.

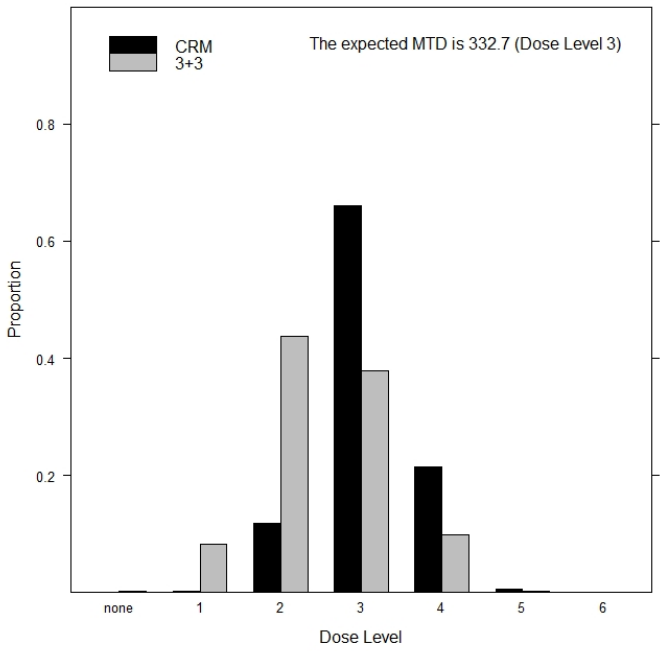

Figure 4.3: Dose Selection for 99% Bioavailability

It is clear from these simulation results that the CRM greatly outperforms the standard $3 + 3$ design in multiple ways. We first consider how often the design selects the expected MTD to recommend for future Phase I trials. The expected

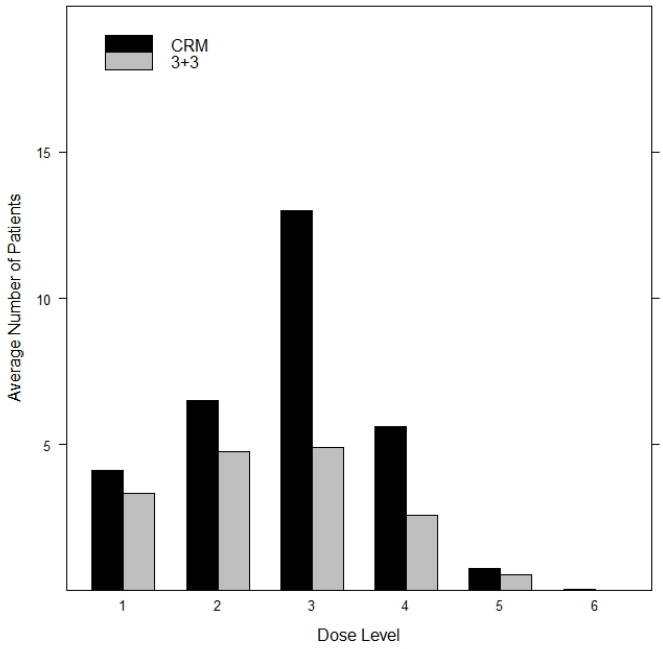

Figure 4.4: Number of Patients Assigned to Each Dose Level for 99% Bioavailability

MTD under 99% bioavailability is the third dose level. Anythng less than this dose level is considered sub-therapeutic. The CRM correctly selected the third dose level 68% of the time, but the $3+3$ design selected the third dose level only 37.8% of the time. In addition, the $3+3$ recommended a sub-therapeutic dose in 53.8% of the trials while the CRM recommended a sub-therapeutic dose in only 12% of the trials.

In the simulation, we also tracked the average number of patients assigned to each of the six dose levels over the 500 trials. The CRM trials in our simulation have a fixed sample size of thirty patients admitted into the trial, while the trials that use the $3+3$ design have an average of 16.122 patients admitted into the trial. In the CRM trials, an average of 12.98 patients were treated at the expected MTD while only an average of 4.9 patients were treated at the expected MTD in the $3+3$ trials. Because of the different number of patients enrolled in these two types of trials, we are unable to make a direct comparison regarding sample sizes. However, we can conclude that the CRM treats a greater proportion of patients at the expected MTD

than the $3+3$ design does. The CRM assigns 43.3% of patients in a hypothetical trial to the expected MTD while the $3+3$ assigns 30.4% of patients to the expected MTD.

The main goal of the Phase I trial is to select the MTD for use in latter phases of the study. A dose that is lower than the MTD is likely to be less effective while a dose that is higher than the MTD will lead to an unacceptably high toxic response in patients. From these results, the CRM outperforms the standard design at 99% bioavailability. The CRM is more likely to recommend the correct dosage, and it treats a greater number of patients at therapeutic doses when compared to the $3+3$.

Thus, as expected, the results for 99% bioavailability are similar to those already in the literature comparing the $3+3$ method and the CRM. We will now compare the two designs when the patient population experiences 65% bioavailability. The results for this scenario can be seen in Figures 4.5 and 4.6.

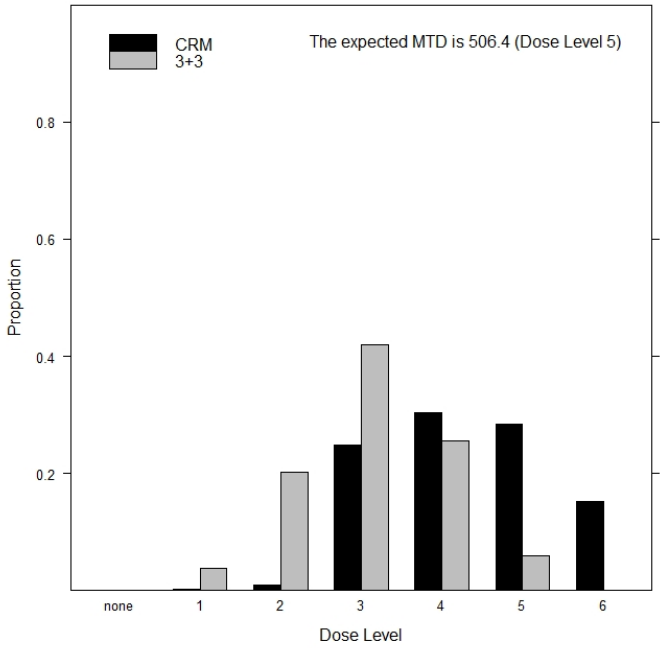

Figure 4.5: Dose Selection for 65% Bioavailability

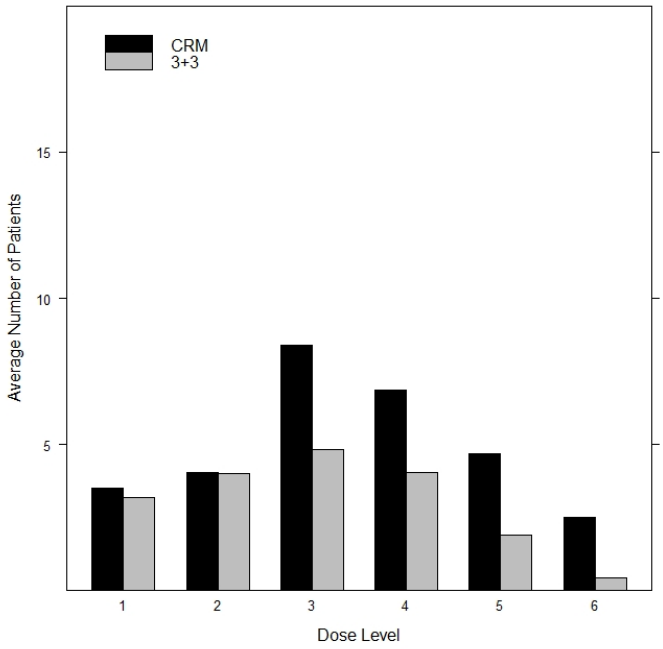

Figure 4.6: Number of Patients Assigned to Each Dose Level for 65% Bioavailability

At this lesser bioavailability, the differences between the $3+3$ method and the CRM become even more pronounced. The CRM is clearly accomodating for the low bioavailability by selecting higher doses than it did when there was 99% bioavailability in the patient population. Out of the 500 simulated trials, the CRM selected the third dose level 24.8% of the time, the fourth dose level 30.4% of the time, and the fifth dose level 28.4% of the time. However, the $3+3$ does not appear to be accomodating the lower bioavailability by consistently selecting higher dose levels. It selects the third dose level 42% of the time, the fourth dose level 25.6% of the time, and the fifth dose level 6% of the time.

From Table 4.3, we know that the expected MTD at 65% bioavailability is 506.4 units with a standard deviation of 70.6. This is equivalent to the fifth dose level. The CRM selected this dose level 28.4% of the time while the $3+3$ only selected this dose level 6% of the time. Based on these results, we can conclude that

the one-parameter logistic model of the CRM outperforms the $3+3$ design when we account for lower bioavailability.

It is interesting to see how the designs are allocating patients to the six dose levels. The $3+3$ design is allocating very few patients to the higher dose levels. It is, in fact, assigning fewer than five patients to the fourth, fifth, and sixth dose levels, which are in the therapeutic range. The CRM, on the other hand, assigns more patients to these therapeutic doses, making it more likely that a therapeutic dose will be recommended for use in the future phases of the trial.

Similar results are seen when we consider bioavailability levels of 90%, 75%, and 35%. The results for these scenarios can be found in Appendix B.

4.4 The CRM Power Model

The logistic models, in particular the one defined in (4.1), are the most commonly used models when the CRM is used in Phase I trials. However, there are other models that should be considered as potential alternatives. Paoletti and Kramar (2009) performed a literature review of 33 Phase I oncology trials that used the CRM. Of those 33 trials, 10 used a logistic model and 6 used a one-parameter power model, as defined below. They claim that for certain dose response scenarios, the power model will outperform the logistic model as well as the $3+3$ design.

In the previous simulation, we saw that the one-parameter logistic model performs well in the presence of low bioavailability levels. We now compare it to the power model, and determine which model performs better when we account for sub-bioavailability.

4.4.1 The Power Model

The one-parameter power model generally takes the form

$$P\left(DLT|x\right) = x^{a}, \tag{4.6}$$

where x is the dose level and a is the parameter of interest. There is no restriction on the range of a, but the dose level must be recoded such that the dose levels are in the $(0, 1)$ interval (Paoletti and Kramar, 2009). In the Bayesian CRM, an exponential prior with parameter one is typically used as the prior for a. We use the same methodology as in the first simulation to simulate values of bioavailability. Therefore, we use the following power model for our simulation:

$$P\left(DLT|x\right) = \left(\gamma_u x\right)^a, \tag{4.7}$$

where γ_u represents the level of bioavailability centered at u as defined in Section 4.3.

4.4.2 Adjusting for Bioavailability

For our simulation, we use the same DLT scenario as in Table 4.2. We are interested in determining how the true value of the MTD changes when the level of bioavailability decreases. Using the one-parameter power model and the beta distributions given in Table 4.1, we consider the dose response curves adjusted for the five levels of bioavailability. As we saw with the one-parameter logistic model, bioavailability has an impact on the dose-response models. In order to determine how much bioavailability is affecting the value of the MTD when the power model is used, we calculate the expected value of the MTD, adjusting for the effect of bioavailability.

The MTD is now a random variable; for a target toxicity rate, p, the solution of

$$p = k\left(\gamma_u x\right)^a$$

is

$$x_{MTD} = \frac{1}{\gamma_u}\sqrt[a]{\frac{p}{k}}, \tag{4.8}$$

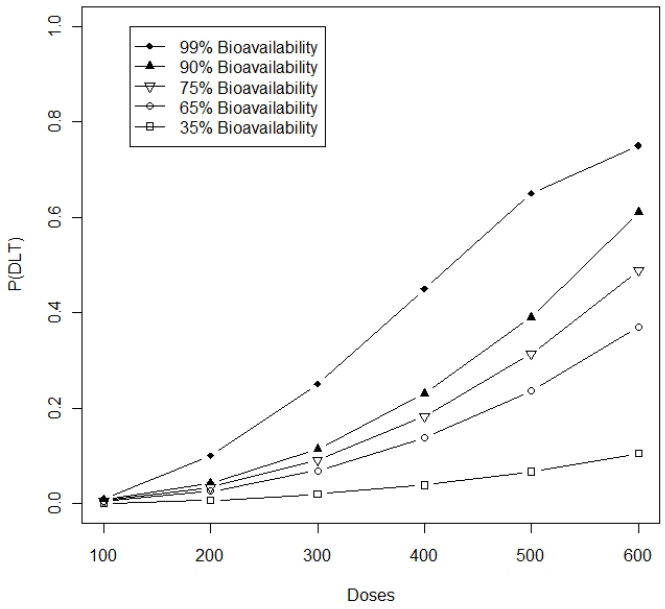

Figure 4.7: Power Dose Response Curves Adjusted for Bioavailability

where k and a are the coefficient and exponent of the true power model, respectively. Therefore,

$$\begin{aligned} \mu_{MTD} &= E\left(x_{MTD}\right) \\ &= E\left(\frac{1}{\gamma_u}\right)\sqrt[a]{\frac{p}{k}} \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} \sigma_{MTD} &= Var\left(X_{MTD}\right) \\ &= \sqrt{Var\left(\frac{1}{\gamma_u}\right)\left(\sqrt[a]{\frac{p}{k}}\right)^2}. \end{aligned} \tag{4.10}$$

Again, we can use (4.5) to calculate the expected value of the MTD under the five bioavailability scenarios. Using (4.9) and (4.10), we calculate the expected value and standard deviations of the MTD at each of the levels of bioavailability.

Table 4.4: Expected Value of the MTD Using the Power Model

Level of Bioavailability	*Distribution*	*Mode*	*SD*	μ_{MTD}	σ_{MTD}
35%	Beta(7.31,12.71)	0.35	0.105	1000.4	355.1
65%	Beta(20.99,11.77)	0.65	0.082	527.0	73.6
75%	Beta(9.63,3.88)	0.75	0.118	480.8	96.9
90%	Beta(5.38,1.49)	0.90	0.147	444.3	121.5
99%	Beta(88.28,1.88)	0.99	0.015	338.9	5.3

4.4.3 Simulation Results

We first compare the performance of the power model to that of the standard 3+3 design. For each of the five levels of bioavailability, the simulation was replicated 500 times. This represents the occurrence of 500 Phase I trials. The results for 99% bioavailability can be seen below in Figures 4.8 and 4.9.

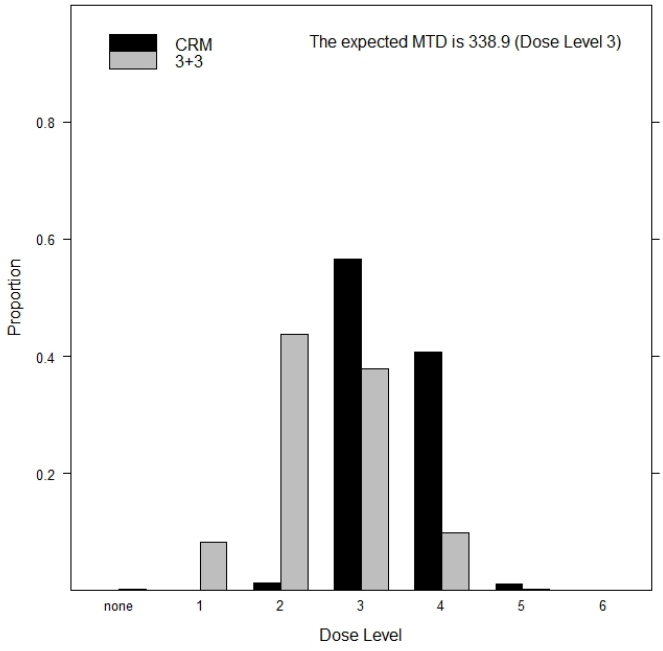

Figure 4.8: Dose Selection for 99% Bioavailability

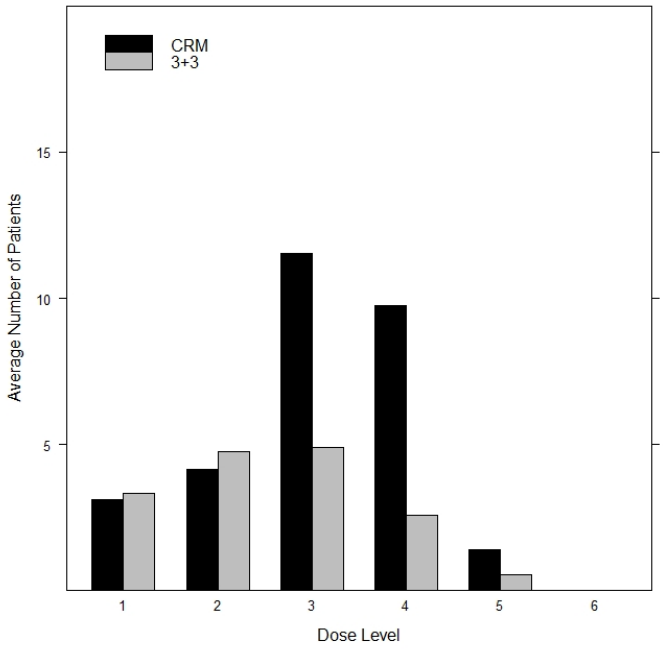

Figure 4.9: Number of Patients Assigned to Each Dose Level for 99% Bioavailability

In Figure 4.8, we consider the accuracy of the two designs. The CRM power model clearly performs better than the $3 + 3$. When the power model is used, the CRM correctly chooses the expected MTD, which is the third dose level, 56.6% of the time. Also, a sub-therapeutic dose level was selected as the MTD only 1.4% of the time. The $3 + 3$ only selected the expected MTD 37.8% of the time, and a sub-therapeutic dose was selected 52.2% of the time.

We also looked at the dose assignments in Figure 4.9. Out of 30 participating patients in each trial, an average of 11.53 were assigned to the correct dose. There were relatively few patients assigned to the lower dose levels. The $3 + 3$, on the other hand, did not assign that many patients to a particular dose. The CRM demonstrates rapid convergence to a particular dose because of the relatively large number of patients assigned to the third dose level. However, the $3 + 3$ design does not demonstrate the same level of convergence.

Again, these results are not surprising for this high level of bioavailability. Given results reported by Paoletti and Kramar (2009) indicating that the power model performs well compared to the logistic model in the CRM, it is expected that the power model-based CRM will outperform the $3 + 3$ method.

We now consider how the power model performs when the level of bioavailability is only 65%. The results for 65% bioavailability can be seen in 4.10 and 4.11.

The CRM power model continues to outperform the standard design at this

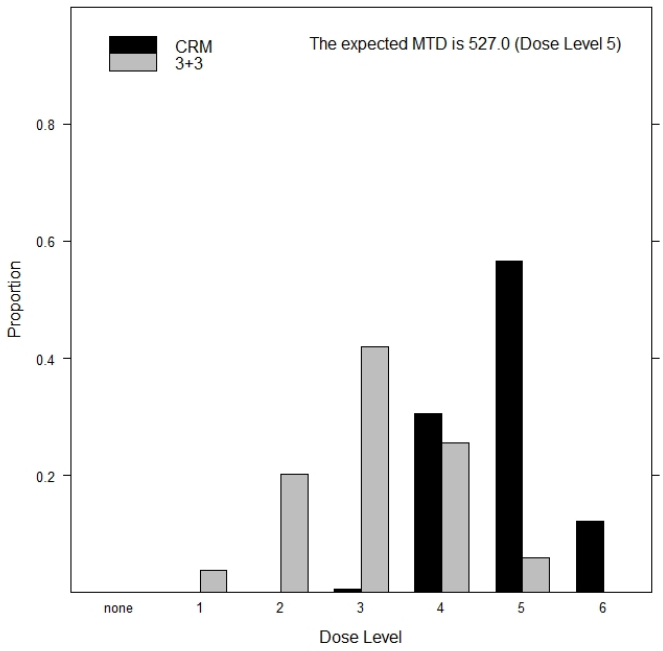

Figure 4.10: Dose Selection for 65% Bioavailability

lower level of bioavailability. In Table 4.4, we found the expected MTD at 65% bioavailability to be 527.0 mg with a standard deviation of 73.6. This is equivalent to the fifth dose level. The CRM power model selects this dose 56.6% of the time while the $3 + 3$ design selects this level only 6% of the time. More importantly, the CRM model selected a sub-therapeutic dose 21.2% of the time while the $3 + 3$ selected a sub-therapeutic dose 94% of the time.

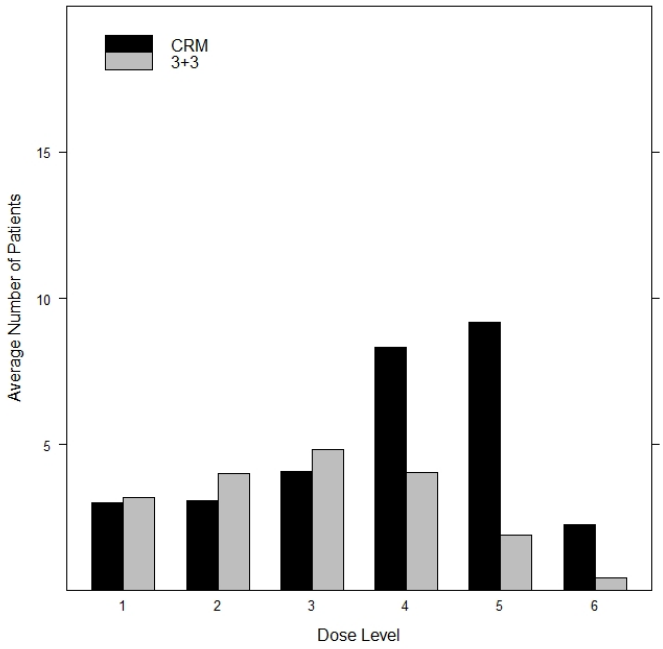

Figure 4.11: Number of Patients Assigned to Each Dose Level for 65% Bioavailability

We now compare the dose level assignments. In the CRM, an average of 9.19 patients were assigned to the fifth dose level in each of the trials. The $3 + 3$ only assigned an average of 1.9 patients to the fifth dose level in each trial. Based on these simulation results, the CRM power model is clearly superior to the standard design when we account for sub-bioavailability.

We conclude, not surprisingly, that the CRM power model outperforms the standard design. However, we are also interested in comparing the power and logistic models under sub-bioavailability. To compare the performance of the two models, we look at the accuracy of the dose selection through the following MTD error plots. A negative value on the horizontal axis indicates that a lower dose than the true MTD was selected while a positive value indicates that a higher dose was selected. A value of zero indicates that the expected MTD was selected. We consider the bioavailability levels of 99% and 65%.

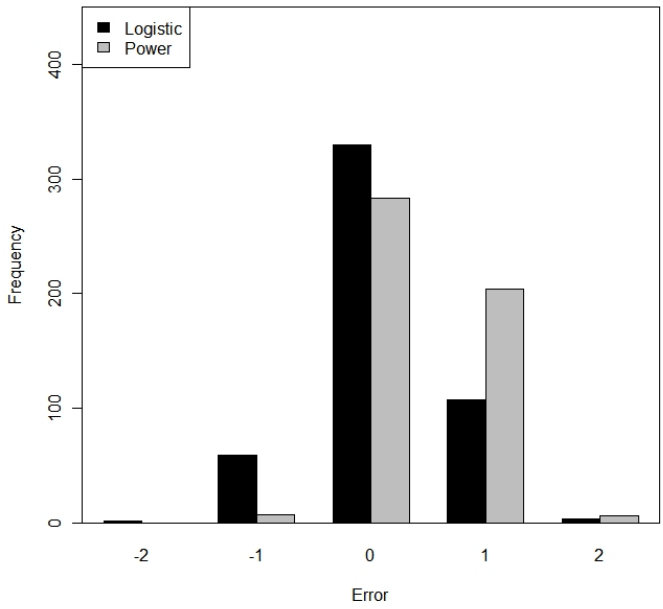

Figure 4.12: MTD Error Plot for 99% Bioavailability

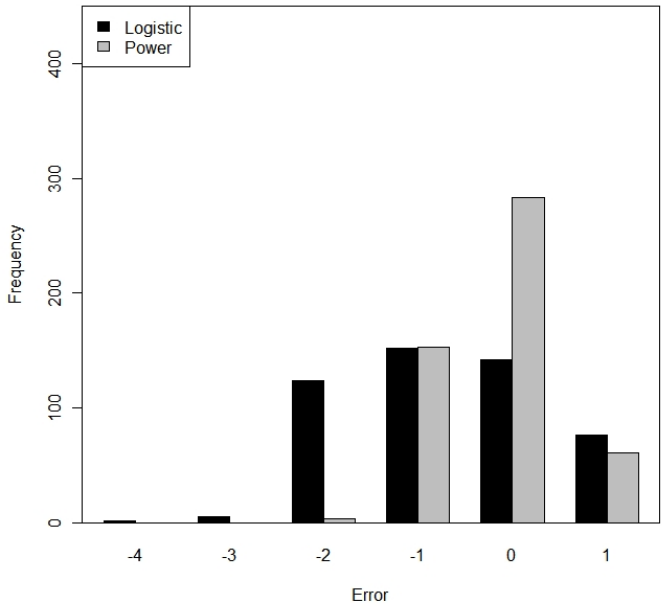

Figure 4.13: MTD Error Plot for 65% Bioavailability

When we have 99% bioavailability in the patient population, the logistic model appears to be performing slightly better than the power model. It is more likely to select the correct dose level to recommend as the MTD. Also, the power model exhibits a greater frequency of recommending a more toxic dose than is appropriate.

However, when we decrease the level of bioavailability to 65%, the power model performs much better than the logistic model. The power model is much more likely to select the correct MTD, and also does not exhibit a high frequency of recommending sub-therapeutic doses.

MTD error plots comparing the logistic and power model for the remaining bioavailability scenarios can be found in Appendix D.

4.5 Bioavailability: A Maximum Absorbable Dose

In Hande et al. (1993), an oncolytic drug was studied at various doses. Prior to the bioavailability study, the physicians thought that the bioavailability would increase as the dose increases. However, the study found that the mean bioavailability for a 100 mg dose was 76% and 48% for a 400 mg dose. Since these results contradicted the phyisicians' prior beliefs, they performed an additional bioavailability study and found that the maximum dose the body can absorb is 250 mg. Instead of a constant rate of bioavailability, we now look at how the CRM responds when there is a maximum absorbable dose.

We use the same doses and probabilities of toxicity as given in Table 4.2. Consider the following model used to simulate the data:

$$logit\left[P(Y=1|x)\right] = 3 + \beta x.$$

We allow the maximum absorbable dose, denoted by η, to be random variable and consider three different distributions for η: $\eta \sim Normal(300, 50)$, $\eta \sim Normal(300, 75)$, and $\eta \sim Normal(500, 50)$. If the assigned dose is greater than the maximum dose, then the simulated toxicity outcome will be based on the maximum absorbed dose.

The typical sample size for a Phase I trial is 30 patients. However, if certain conditions are met, the trial can be stopped early. How does the CRM performs if it is "aware" of the maximum absorbable dose and is allowed to stop early?

For our simulation, we allow the CRM to stop early if six patients have been treated at the same dose, and the seventh patient is assigned that same dose. This is the typical stopping rule employed in Phase I clinical trials (Ahn, 1998). For each maximum absorbable dose, we compare two properties. First, we want to know if there is a significant reduction in sample size if the stopping rule is employed. Second, we want to know if the MTD is still correctly selected if the sample size is reduced.

We first consider the case where $\eta \sim N(300, 75)$.

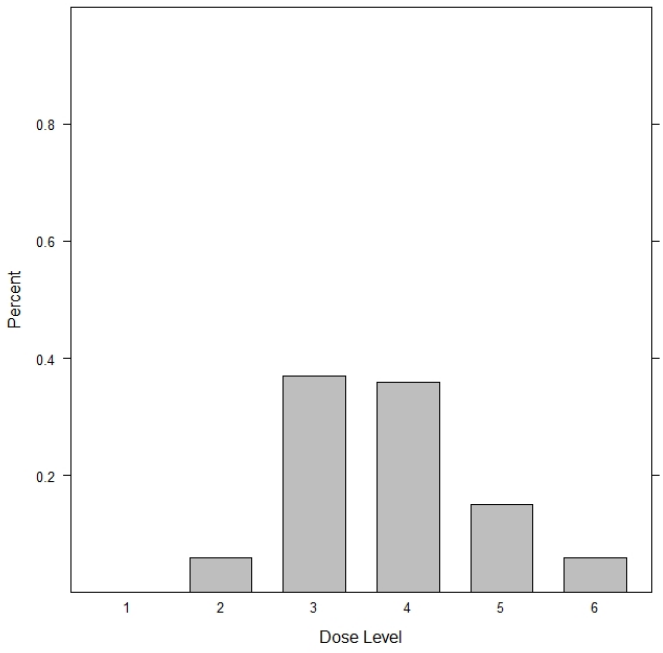

Figure 4.14: Dose Selection with 30 Patients - $\eta \sim N(300, 75)$

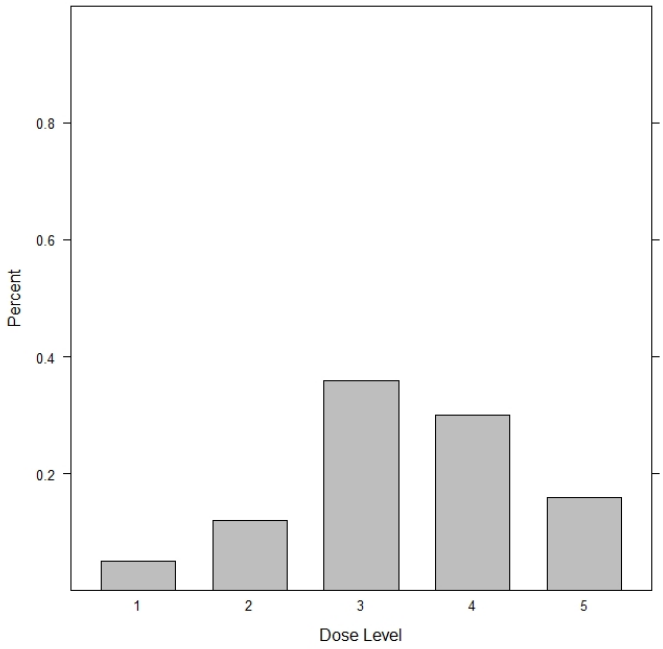

Figure 4.15: Dose Selection with an Average of 14.79 Patients - $\eta \sim Normal(300, 75)$

The CRM appears to be adjusting for the maximum absorbable dose by recommending slightly higher dose levels. However, even in the cases where the stopping rule is employed, it still recommends the correct MTD the majority of the time.The remaining results for this simulation can be found in Appendix E.

4.6 Discussion

In this chapter, we have looked at the effect of bioavailability on the performance of the CRM and the $3 + 3$ design. It is clear that accounting for the effect of bioavailability can have an important impact on the value of the MTD that is recommended for use in future trials. Based on our results and because of the importance of this recommended dose, performing bioavailability studies prior to the onset of the Phase I trial would benefit the effectiveness and efficiency of the trial.

We also found that the power model-based CRM outperforms the logistic model-based CRM when the bioavailability is low. We plan on further investigating

the reasons for this performance.

When the stopping rule is employed under the maximum absorbable dose scenario, the sample size for the trial is significantly reduced to between 13-15 patients. In a Phase I trial, the goal is to reach the MTD as quickly as possible in order to avoid treating patients at either ineffective or toxic doses. While we do not know what precisely what is causing this reduced sample size, it is possible that if the bioavailability of the drug has been studied prior to the onset of the trial, this information can be incorporated into the trial design and the MTD can be selected much more quickly.

APPENDICES

APPENDIX A

Censored Binomial Simulation Results

Table A.1: Simulation Results for $\theta = 0.05$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.404	0.093	0.324	0.119	0.285	0.129
Posterior Mean	0.436	0.063	0.365	0.097	0.322	0.117
Posterior SD	0.206	0.02	0.182	0.041	0.164	0.057
Likelihood Interval Lower Bound	0.032	0.024	0.024	0.024	0.02	0.02
Likelihood Interval Upper Bound	0.870	0.088	0.743	0.171	0.658	0.225
Bayesian Interval Lower Bound	0.079	0.024	0.063	0.027	0.052	0.025
Bayesian Interval Upper Bound	0.832	0.071	0.722	0.142	0.637	0.196
Likelihood Interval Width	0.839	0.083	0.719	0.167	0.638	0.222
Bayesian Interval Width	0.753	0.059	0.659	0.128	0.585	0.182

Table A.2: Simulation Results for $\theta = 0.05$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.361	0.084	0.188	0.055	0.108	0.04
Posterior Mean	0.408	0.057	0.248	0.045	0.148	0.035
Posterior SD	0.195	0.012	0.130	0.01	0.079	0.008
Likelihood Interval Lower Bound	0.034	0.031	0.021	0.022	0.016	0.017
Likelihood Interval Upper Bound	0.825	0.069	0.52	0.065	0.306	0.052
Bayesian Interval Lower Bound	0.077	0.03	0.046	0.023	0.03	0.018
Bayesian Interval Upper Bound	0.796	0.056	0.538	0.055	0.332	0.047
Likelihood Interval Width	0.792	0.056	0.499	0.049	0.29	0.037
Bayesian Interval Width	0.719	0.038	0.492	0.036	0.302	0.03

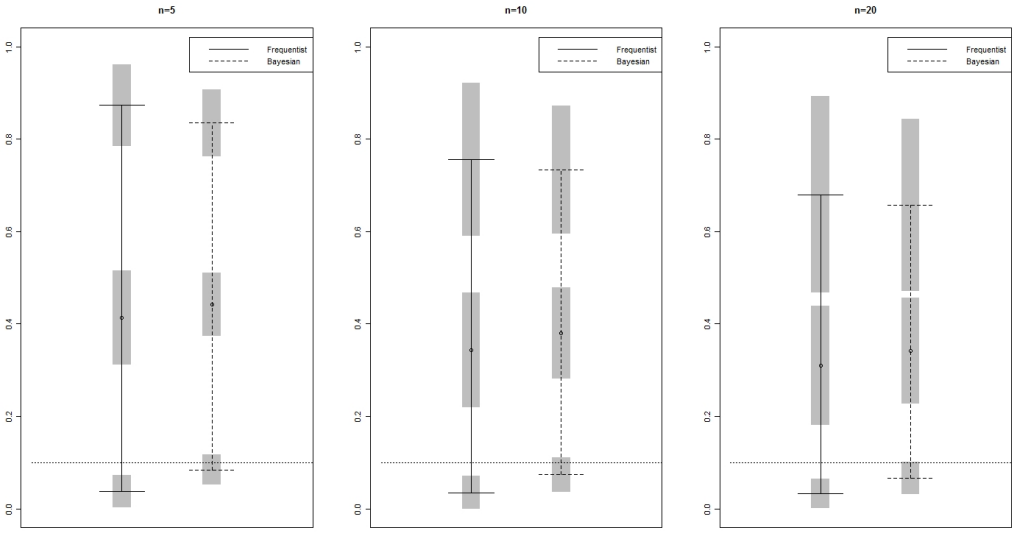

Figure A.1: Simulation Results for $\theta = 0.10$ - Unrestricted Censoring Intervals

Table A.3: Simulation Results for $\theta = 0.10$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.413	0.102	0.344	0.124	0.31	0.129
Posterior Mean	0.442	0.068	0.38	0.099	0.342	0.115
Posterior SD	0.205	0.02	0.182	0.039	0.166	0.054
Likelihood Interval Lower Bound	0.038	0.035	0.035	0.036	0.032	0.032
Likelihood Interval Upper Bound	0.873	0.089	0.756	0.166	0.68	0.212
Bayesian Interval Lower Bound	0.085	0.033	0.074	0.037	0.066	0.035
Bayesian Interval Upper Bound	0.835	0.072	0.734	0.139	0.657	0.186
Likelihood Interval Width	0.835	0.082	0.721	0.16	0.647	0.21
Bayesian Interval Width	0.75	0.059	0.66	0.122	0.591	0.171

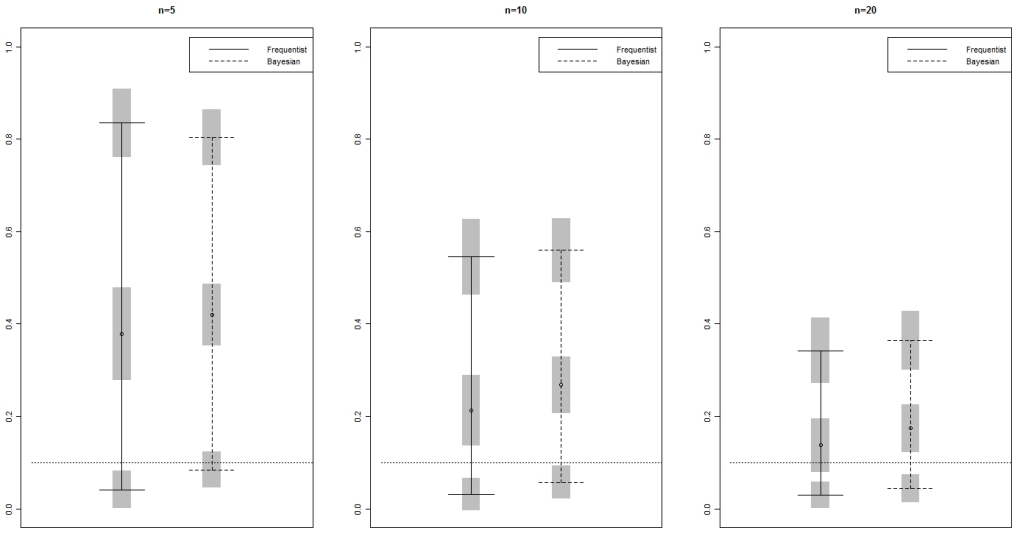

Figure A.2: Simulation Results for $\theta = 0.10$ - Restricted Censoring Intervals

Table A.4: Simulation Results for $\theta = 0.10$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.379	0.10	0.213	0.076	0.137	0.058
Posterior Mean	0.42	0.067	0.268	0.062	0.174	0.052
Posterior SD	0.195	0.012	0.133	0.01	0.084	0.009
Likelihood Interval Lower Bound	0.042	0.041	0.032	0.035	0.03	0.029
Likelihood Interval Upper Bound	0.835	0.074	0.545	0.082	0.342	0.071
Bayesian Interval Lower Bound	0.084	0.039	0.057	0.036	0.045	0.03
Bayesian Interval Upper Bound	0.804	0.060	0.56	0.069	0.364	0.064
Likelihood Interval Width	0.793	0.055	0.513	0.053	0.312	0.044
Bayesian Interval Width	0.72	0.038	0.502	0.039	0.32	0.035

Table A.5: Simulation Results for $\theta = 0.25$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.447	0.115	0.40	0.131	0.382	0.131
Posterior Mean	0.465	0.076	0.422	0.101	0.4	0.11
Posterior SD	0.204	0.018	0.182	0.035	0.167	0.047
Likelihood Interval Lower Bound	0.057	0.054	0.068	0.064	0.078	0.065
Likelihood Interval Upper Bound	0.889	0.083	0.796	0.147	0.738	0.181
Bayesian Interval Lower Bound	0.102	0.049	0.106	0.006	0.112	0.064
Bayesian Interval Upper Bound	0.85	0.069	0.769	0.124	0.712	0.16
Likelihood Interval Width	0.833	0.077	0.728	0.139	0.659	0.179
Bayesian Interval Width	0.748	0.055	0.664	0.107	0.6	0.146

Table A.6: Simulation Results for $\theta = 0.25$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.425	0.128	0.305	0.126	0.256	0.099
Posterior Mean	0.45	0.085	0.342	0.102	0.28	0.089
Posterior SD	0.195	0.011	0.14	0.01	0.097	0.009
Likelihood Interval Lower Bound	0.064	0.059	0.079	0.073	0.099	0.065
Likelihood Interval Upper Bound	0.859	0.08	0.632	0.115	0.475	0.106
Bayesian Interval Lower Bound	0.105	0.054	0.104	0.07	0.114	0.064
Bayesian Interval Upper Bound	0.826	0.068	0.635	0.1	0.486	0.098
Likelihood Interval Width	0.795	0.052	0.553	0.053	0.376	0.044
Bayesian Interval Width	0.721	0.036	0.530	0.039	0.372	0.037

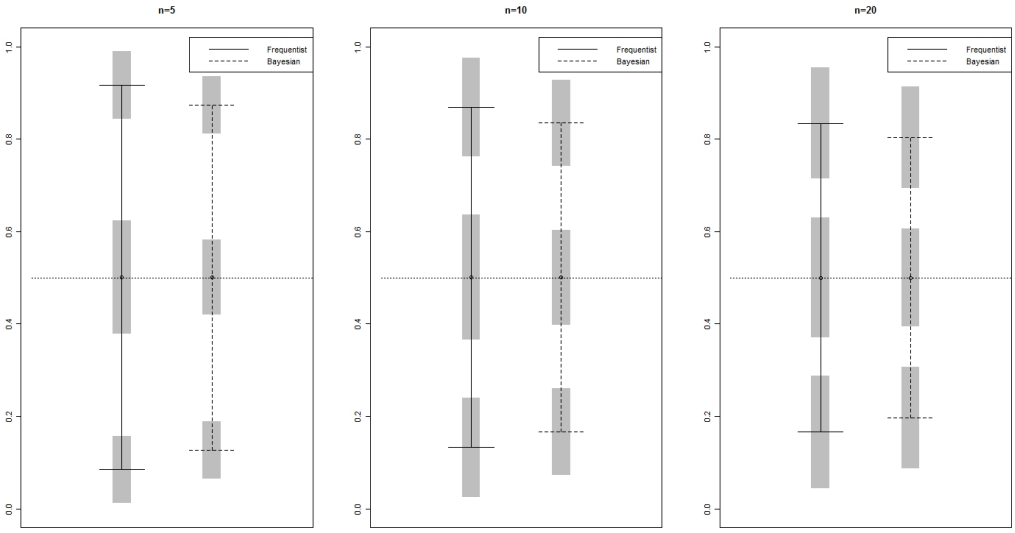

Figure A.3: Simulation Results for $\theta = 0.50$ - Unrestricted Censoring Intervals

Table A.7: Simulation Results for $\theta = 0.50$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.501	0.122	0.501	0.136	0.5	0.13
Posterior Mean	0.501	0.081	0.05	0.103	0.05	0.106
Posterior SD	0.204	0.018	0.182	0.030	0.168	0.04
Likelihood Interval Lower Bound	0.085	0.073	0.133	0.107	0.166	0.121
Likelihood Interval Upper Bound	0.917	0.073	0.868	0.107	0.834	0.120
Bayesian Interval Lower Bound	0.127	0.062	0.166	0.094	0.197	0.10
Bayesian Interval Upper Bound	0.874	0.062	0.835	0.094	0.803	0.109
Likelihood Interval Width	0.831	0.074	0.735	0.119	0.668	0.15
Bayesian Interval Width	0.747	0.053	0.669	0.092	0.606	0.122

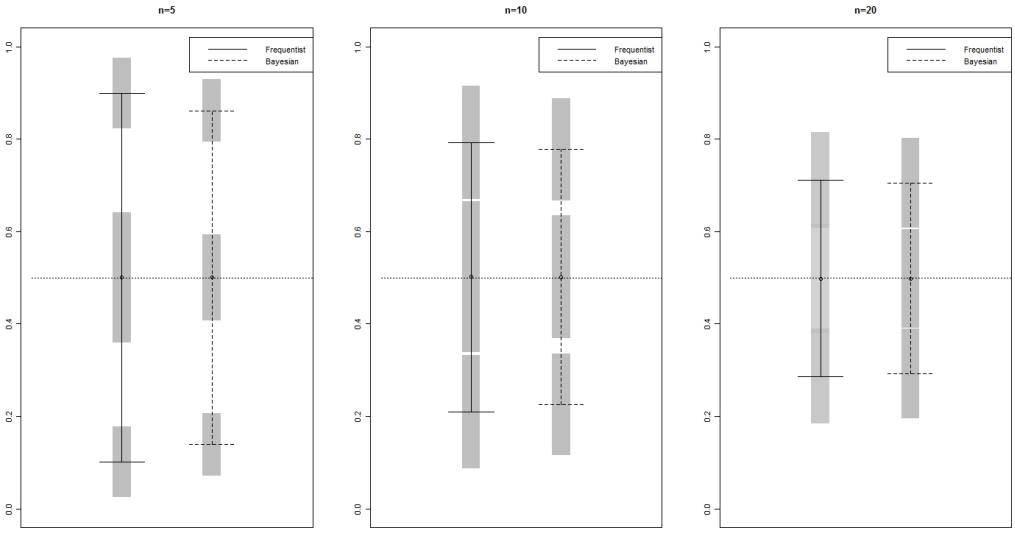

Figure A.4: Simulation Results for $\theta = 0.50$ - Restricted Censoring Intervals

Table A.8: Simulation Results for $\theta = 0.50$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.5	0.141	0.502	0.164	0.498	0.118
Posterior Mean	0.5	0.094	0.502	0.133	0.498	0.107
Posterior SD	0.195	0.011	0.145	0.007	0.107	0.004
Likelihood Interval Lower Bound	0.101	0.076	0.210	0.122	0.287	0.103
Likelihood Interval Upper Bound	0.899	0.076	0.793	0.122	0.711	0.104
Bayesian Interval Lower Bound	0.139	0.067	0.225	0.11	0.293	0.097
Bayesian Interval Upper Bound	0.861	0.067	0.777	0.11	0.705	0.097
Likelihood Interval Width	0.798	0.05	0.583	0.039	0.424	0.018
Bayesian Interval Width	0.722	0.035	0.552	0.029	0.412	0.015

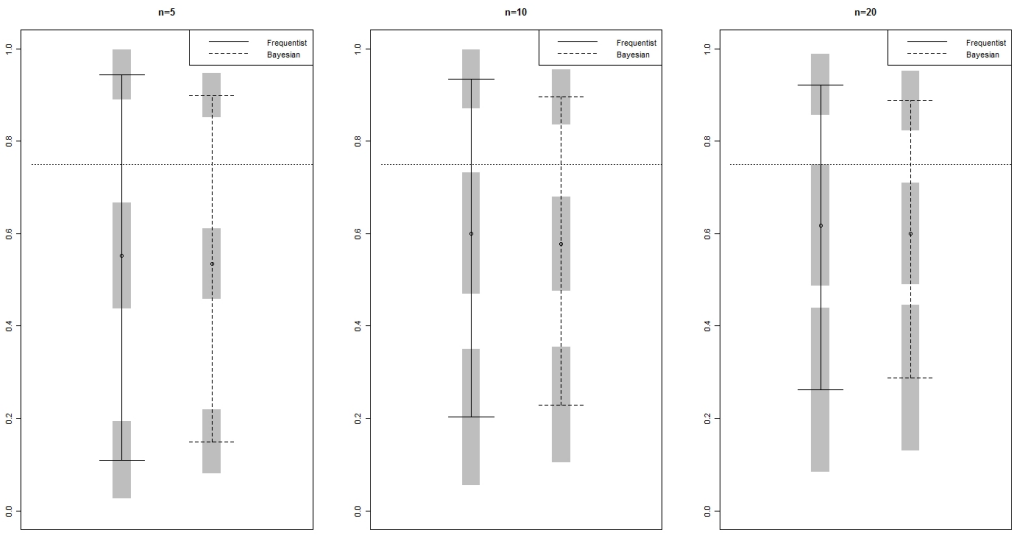

Figure A.5: Simulation Results for $\theta = 0.75$ - Unrestricted Censoring Intervals

Table A.9: Simulation Results for $\theta = 0.75$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.552	0.114	0.6	0.131	0.618	0.131
Posterior Mean	0.535	0.076	0.578	0.102	0.6	0.11
Posterior SD	0.205	0.019	0.182	0.035	0.167	0.046
Likelihood Interval Lower Bound	0.11	0.083	0.203	0.147	0.262	0.178
Likelihood Interval Upper Bound	0.944	0.054	0.933	0.064	0.922	0.066
Bayesian Interval Lower Bound	0.149	0.069	0.23	0.125	0.287	0.158
Bayesian Interval Upper Bound	0.899	0.048	0.895	0.06	0.887	0.064
Likelihood Interval Width	0.834	0.077	0.731	0.139	0.66	0.176
Bayesian Interval Width	0.749	0.056	0.666	0.107	0.6	0.143

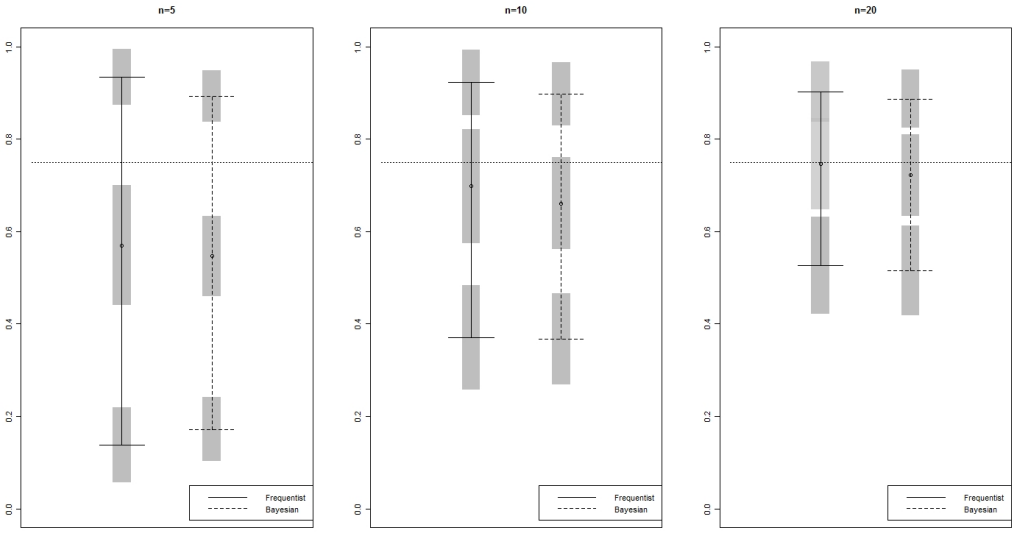

Figure A.6: Simulation Results for $\theta = 0.75$ - Restricted Censoring Intervals

Table A.10: Simulation Results for $\theta = 0.75$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.57	0.13	0.698	0.123	0.746	0.099
Posterior Mean	0.547	0.086	0.66	0.1	0.722	0.088
Posterior SD	0.195	0.011	0.14	0.01	0.096	0.009
Likelihood Interval Lower Bound	0.138	0.081	0.37	0.113	0.527	0.106
Likelihood Interval Upper Bound	0.934	0.061	0.922	0.071	0.902	0.065
Bayesian Interval Lower Bound	0.172	0.069	0.367	0.098	0.516	0.097
Bayesian Interval Upper Bound	0.893	0.056	0.897	0.068	0.887	0.064
Likelihood Interval Width	0.796	0.052	0.552	0.053	0.375	0.044
Bayesian Interval Width	0.721	0.036	0.53	0.039	0.371	0.036

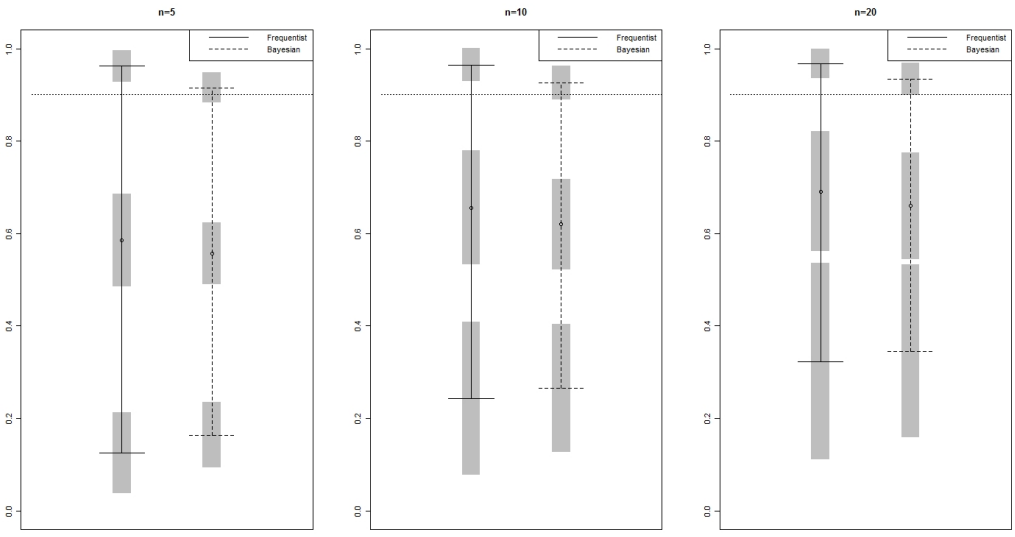

Figure A.7: Simulation Results for $\theta = 0.90$ - Unrestricted Censoring Intervals

Table A.11: Simulation Results for $\theta = 0.90$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.587	0.1	0.656	0.123	0.691	0.130
Posterior Mean	0.557	0.067	0.62	0.098	0.66	0.115
Posterior SD	0.206	0.019	0.182	0.039	0.165	0.054
Likelihood Interval Lower Bound	0.126	0.087	0.243	0.165	0.323	0.213
Likelihood Interval Upper Bound	0.962	0.034	0.965	0.036	0.967	0.032
Bayesian Interval Lower Bound	0.164	0.071	0.265	0.138	0.345	0.187
Bayesian Interval Upper Bound	0.915	0.032	0.926	0.037	0.934	0.036
Likelihood Interval Width	0.836	0.081	0.723	0.16	0.644	0.21
Bayesian Interval Width	0.751	0.058	0.661	0.122	0.588	0.172

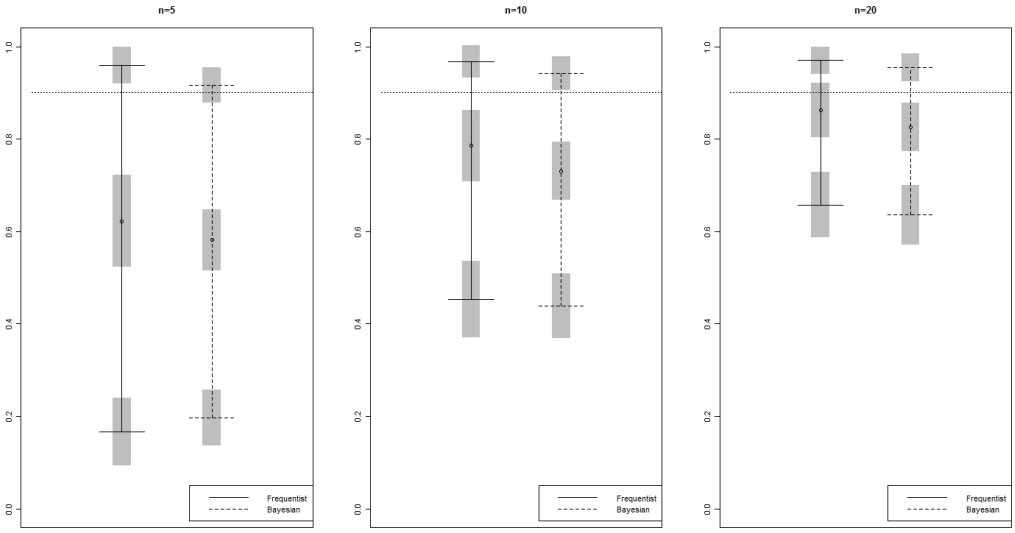

Figure A.8: Simulation Results for $\theta = 0.90$ - Restricted Censoring Intervals

Table A.12: Simulation Results for $\theta = 0.90$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.623	0.099	0.785	0.077	0.862	0.059
Posterior Mean	0.582	0.067	0.731	0.062	0.826	0.052
Posterior SD	0.195	0.012	0.133	0.011	0.084	0.009
Likelihood Interval Lower Bound	0.166	0.073	0.454	0.082	0.658	0.071
Likelihood Interval Upper Bound	0.959	0.04	0.968	0.035	0.97	0.03
Bayesian Interval Lower Bound	0.197	0.06	0.439	0.07	0.636	0.064
Bayesian Interval Upper Bound	0.916	0.038	0.942	0.036	0.955	0.031
Likelihood Interval Width	0.793	0.055	0.514	0.054	0.312	0.044
Bayesian Interval Width	0.72	0.038	0.503	0.039	0.319	0.036

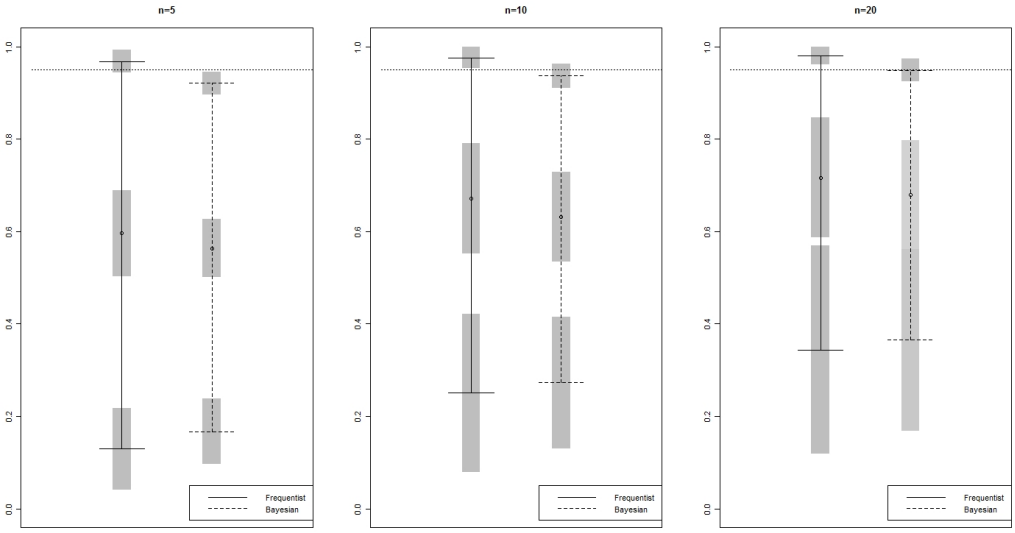

Figure A.9: Simulation Results for $\theta = 0.95$ - Unrestricted Censoring Intervals

Table A.13: Simulation Results for $\theta = 0.95$ - Unrestricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.596	0.093	0.672	0.12	0.716	0.129
Posterior Mean	0.564	0.063	0.631	0.097	0.679	0.117
Posterior SD	0.206	0.02	0.183	0.041	0.164	0.057
Likelihood Interval Lower Bound	0.129	0.088	0.25	0.171	0.344	0.225
Likelihood Interval Upper Bound	0.968	0.024	0.976	0.023	0.981	0.019
Bayesian Interval Lower Bound	0.167	0.071	0.273	0.142	0.365	0.197
Bayesian Interval Upper Bound	0.921	0.024	0.937	0.026	0.948	0.025
Likelihood Interval Width	0.839	0.083	0.726	0.167	0.637	0.223
Bayesian Interval Width	0.753	0.059	0.664	0.128	0.583	0.183

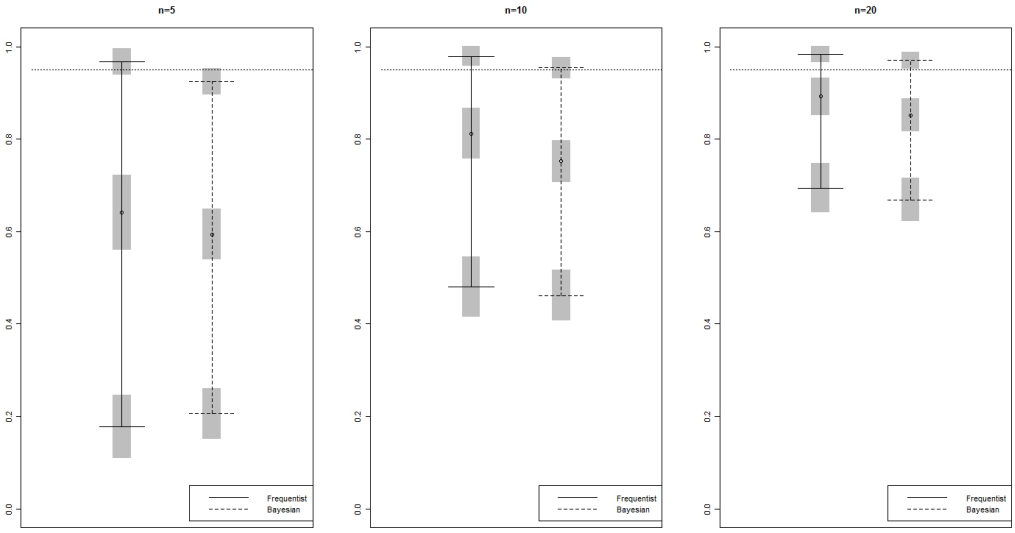

Figure A.10: Simulation Results for $\theta = 0.95$ - Restricted Censoring Intervals

Table A.14: Simulation Results for $\theta = 0.95$ - Restricted Censoring Intervals

	$n = 5$		$n = 10$		$n = 20$	
Estimate	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
MLE	0.642	0.081	0.812	0.055	0.892	0.04
Posterior Mean	0.594	0.055	0.752	0.045	0.852	0.036
Posterior SD	0.194	0.012	0.131	0.01	0.079	0.008
Likelihood Interval Lower Bound	0.177	0.068	0.48	0.065	0.694	0.053
Likelihood Interval Upper Bound	0.967	0.029	0.979	0.022	0.948	0.017
Bayesian Interval Lower Bound	0.206	0.055	0.462	0.055	0.668	0.047
Bayesian Interval Upper Bound	0.924	0.028	0.954	0.023	0.97	0.019
Likelihood Interval Width	0.79	0.056	0.499	0.049	0.29	0.037
Bayesian Interval Width	0.718	0.038	0.493	0.036	0.301	0.03

APPENDIX B

Bioavailability Simulation Results for the One Parameter Logistic Model

B.1 90% Bioavailability

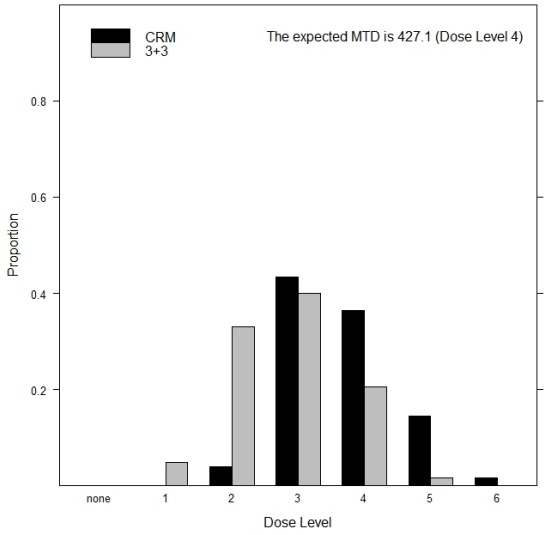

Figure B.1: Dose Selection for 90% Bioavailability

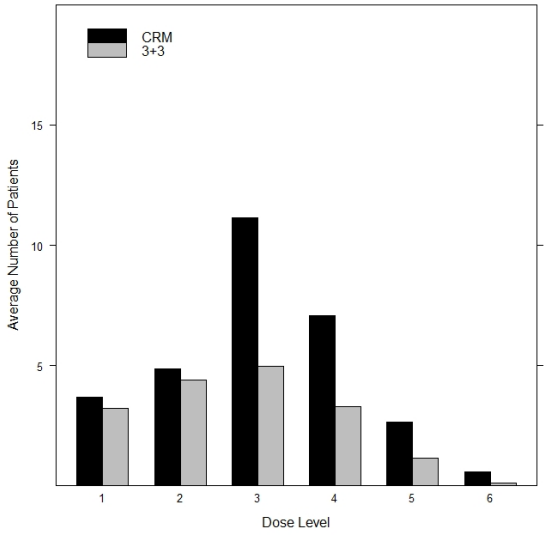

Figure B.2: Number of Patients Assigned to Each Dose Level for 90% Bioavailability

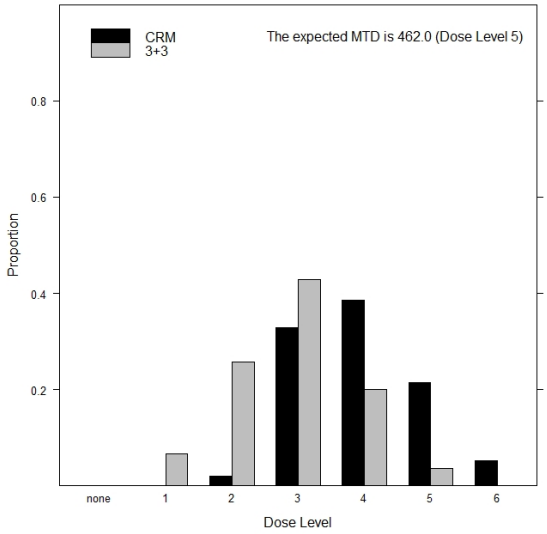

Figure B.3: Dose Selection for 75% Bioavailability

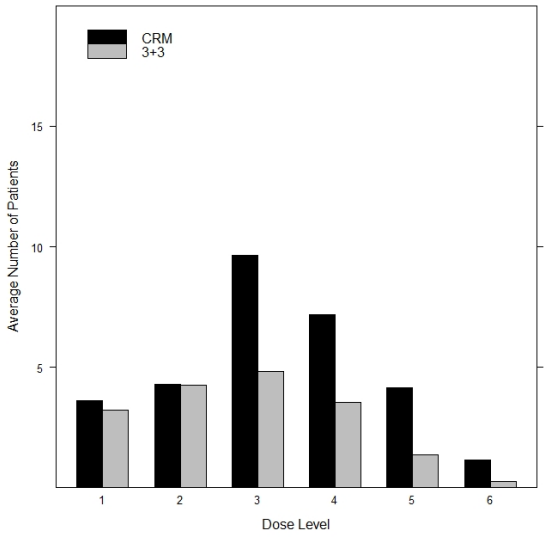

Figure B.4: Number of Patients Assigned to Each Dose Level for 75% Bioavailability

B.3 35% Bioavailability

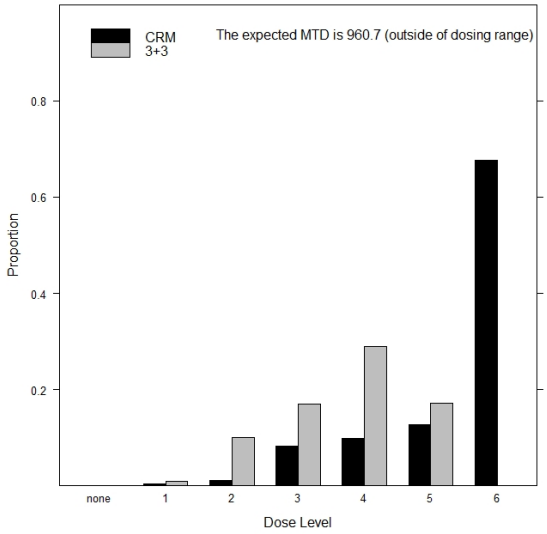

Figure B.5: Dose Selection for 35% Bioavailability

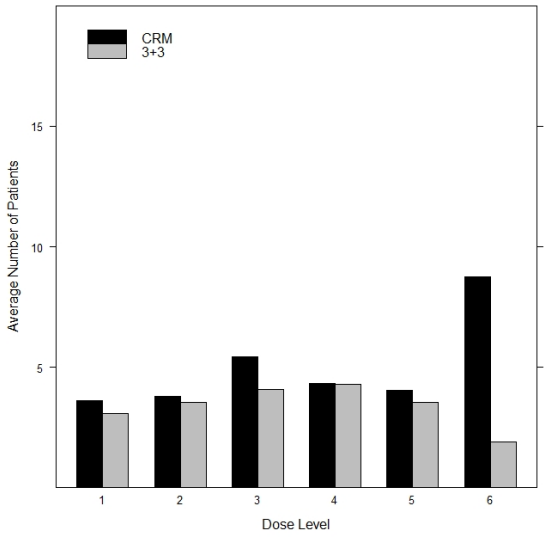

Figure B.6: Number of Patients Assigned to Each Dose Level for 35% Bioavailability

APPENDIX C

Bioavailability Simulation Results for the One Parameter Power Model

C.1 90% Bioavailability

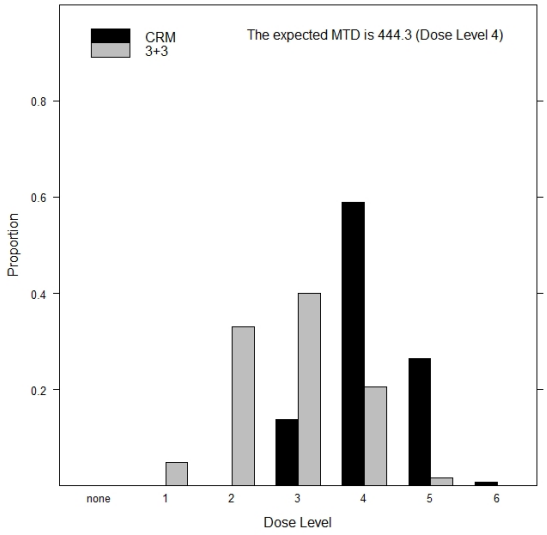

Figure C.1: Dose Selection for 90% Bioavailability

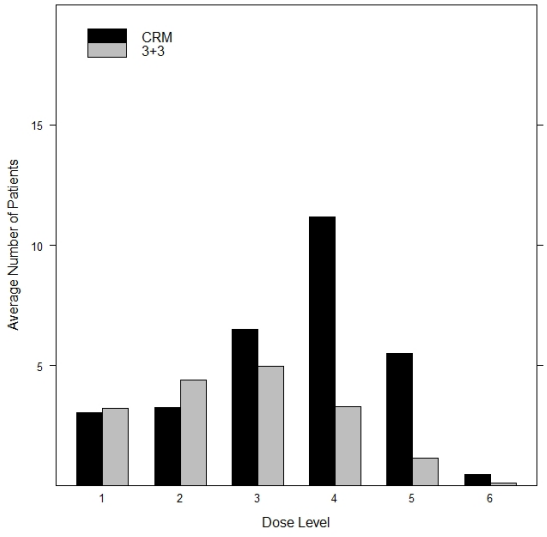

Figure C.2: Number of Patients Assigned to Each Dose Level for 90% Bioavailability

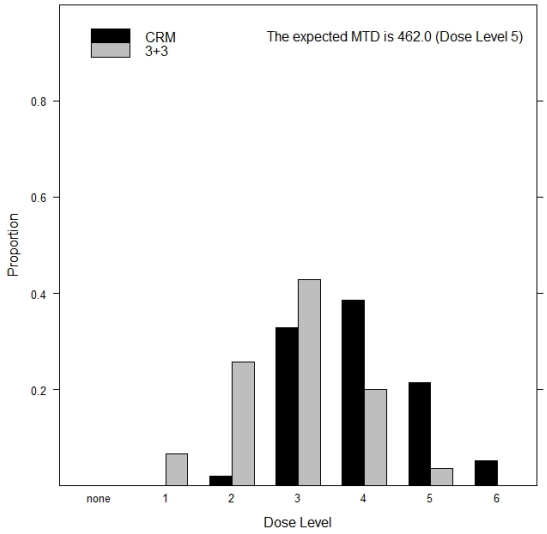

Figure C.3: Dose Selection for 75% Bioavailability

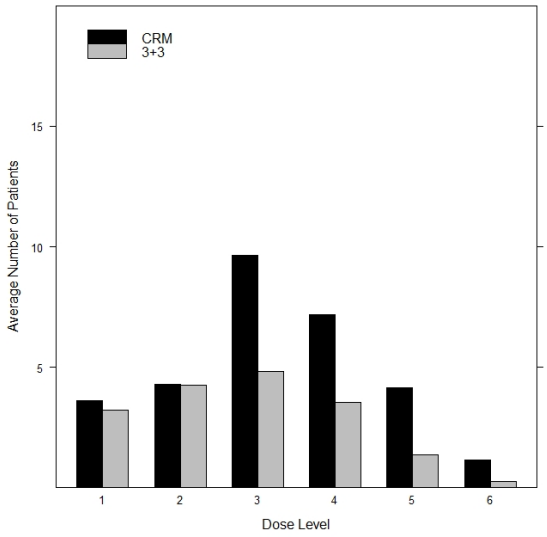

Figure C.4: Number of Patients Assigned to Each Dose Level for 75% Bioavailability

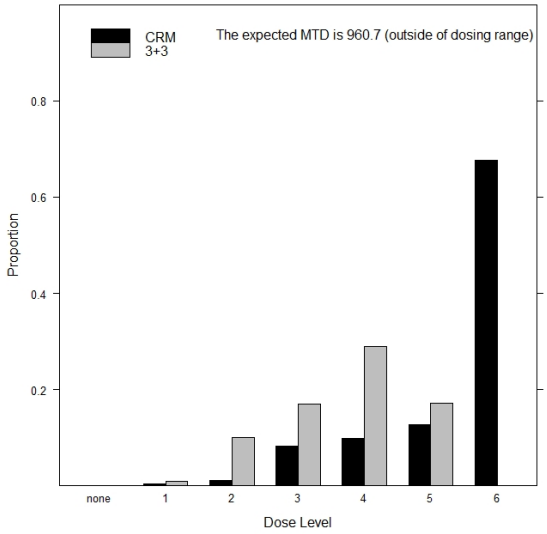

Figure C.5: Dose Selection for 35% Bioavailability

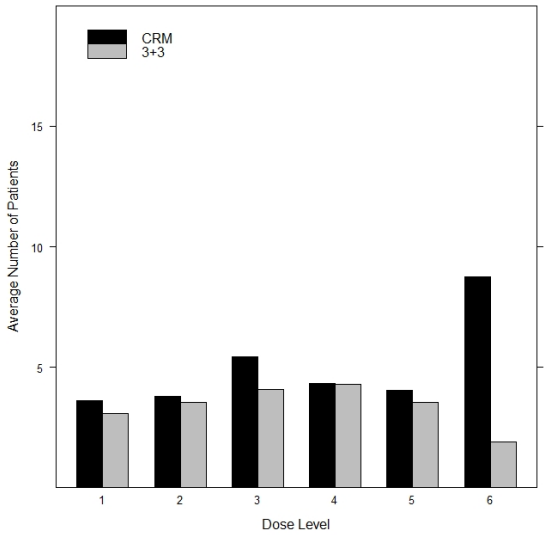

Figure C.6: Number of Patients Assigned to Each Dose Level for 35% Bioavailability

APPENDIX D

Error Plots Comparing the One Parameter Logistic and Power Models

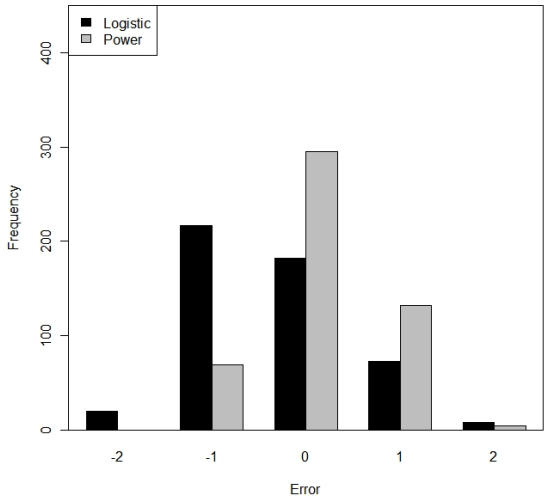

Figure D.1: MTD Error Plot - 90% Bioavailability

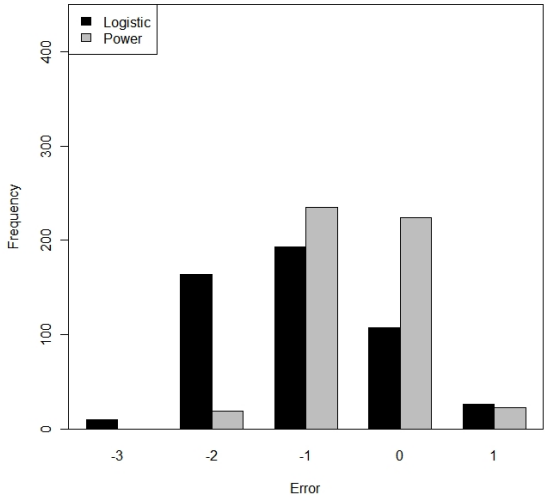

Figure D.2: MTD Error Plot - 75% Bioavailability

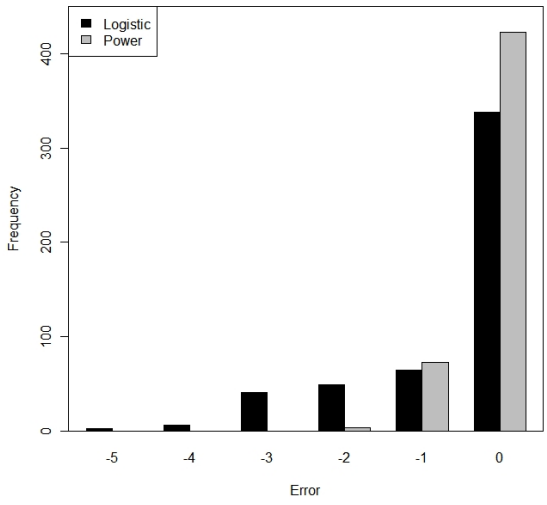

Figure D.3: MTD Error Plot - 35% Bioavailability

APPENDIX E

Results from the Maximum Absorbable Dose Scenario

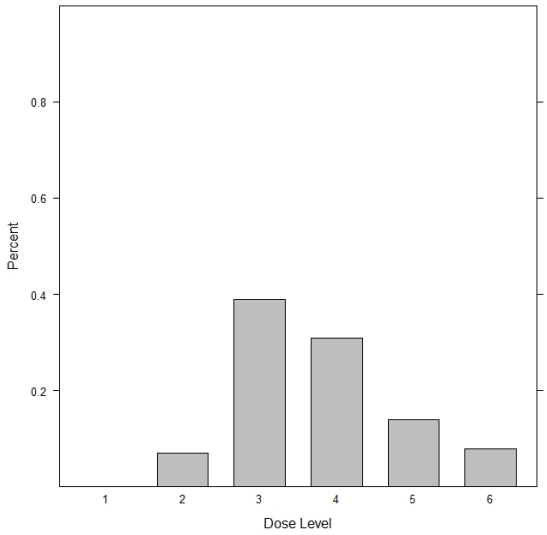

Figure E.1: Dose Selection with 30 Patients - $\eta \sim Normal(300, 50)$

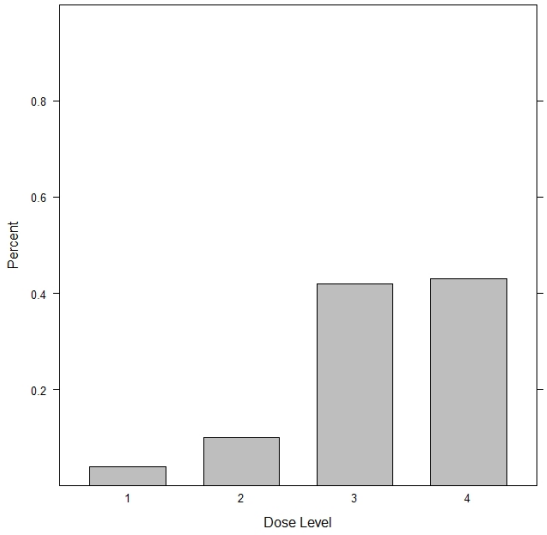

Figure E.2: Dose Selection with an Average of 13.89 Patients - $\eta \sim Normal(300, 50)$

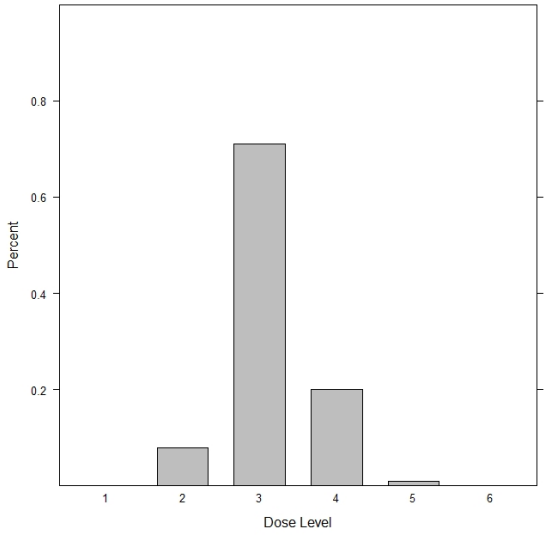

Figure E.3: Dose Selection with 30 Patients - $\eta \sim Normal(500, 50)$

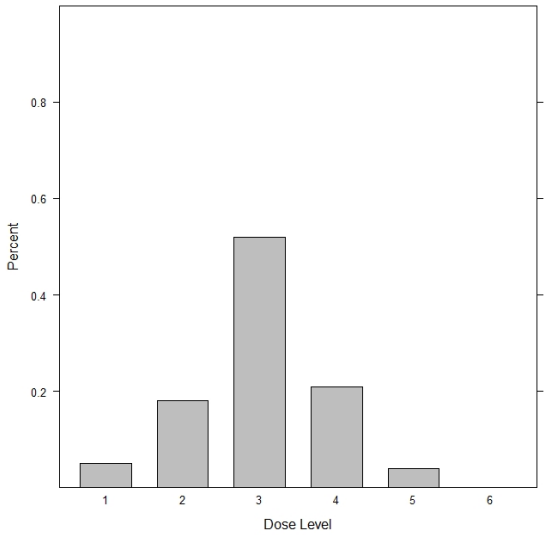

Figure E.4: Dose Selection with an Average of 14.45 Patients - $\eta \sim Normal(500, 50)$

BIBLIOGRAPHY

Ahn, C. (1998), "An Evaluation of Phase I Cancer Clinical Trial Designs," *Statistics in Medicine*, 17, 1537–1549.

David, H. and Nagaraja, H. (2003), *Order Statistics*, Hoboken, NJ: John Wiley and Sons, Inc.

Eisenhauer, E., Twelves, C., and Buyse, M. (2006), *Phase I Cancer Clinical Trials: A Practical Guide*, Oxford University Press.

Fabre, E., Chevret, S., Piechaud, J.-F., Rey, E., Vauzelle-Kervroedan, F., D'Athis, P., Olive, G., and Pons, G. (1998), "An Approach for Dose Finding of Drugs in Infants: Sedation by Midazolam Studied Using the Continual Reassessment Method," *British Journal of Clinical Pharmacology*, 46, 395–401.

Frey, J. and Marrero, O. (2008), "A Surprising MLE for Interval-Censored Binomial Data," *American Statistician*, 62, 135–137.

Garrett-Mayer, E. (2006), "The Continual Reassessment Method for Dose-Finding Studies: A Tutorial," *Clinical Trials*, 3, 57–71.

Gupta, A. K. and Nadarajah, S. (eds.) (2004), *Handbook of Beta Distribution and Its Applications*, New York, NY: Marcel Dekker, Inc.

Hande, K. R., Krozely, M. G., Greco, F. A., Hainsworth, J. D., and Johnson, D. H. (1993), "Bioavailability of Low-Dose Oral Etoposide," *Journal of Clinical Oncology*, 11, 374–377.

Iasonos, A., Wilton, A. S., Seshan, V. E., and Spriggs, D. R. (2008), "A Comprehensive Comparison of the Continual Reassessment Method to the Standard 3+3 Dose Escalation Scheme in Phase I Dose-Finding Studies," *Clinical Trials*, 5, 465–477.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, New York, NY: Springer Science+Business Media, LLC.

O'Quigley, J. and Chevret, S. (1991), "Methods for Dose Finding Studies in Cancer Clinical Trials: A Review and Results of a Monte Carlo Study," *Statistics in Medicine*, 10, 1647–1664.

O'Quigley, J., Pepe, M., and Fisher, L. (1990), "Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer," *Biometrics*, 46, 33–48.

Paoletti, X. and Kramar, A. (2009), "A Comparison of Model Choices for the Continual Reassessment Method in Phase I Cancer Trials," *Statistics in Medicine*, 28, 3012–3028.

Pawitan, Y. (2001), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press.

Potter, D. M. (2006), "Phase I Studies of Chemotherapeutic Agents in Cancer Patients: A Review of the Designs," *Journal of Biopharmaceutical Statistics*, 16, 579–604.

Shardell, M., Scharfstein, D. O., and Bozzette, S. A. (2007), "Survival Curve Estimation for Informatively Coarsened Discrete Event-Time Data:," *Statistics in Medicine*, 26, 2184–2202.

Shu, J. and O'Quigley, J. (2008), "Commentary: Dose-Escalation Designs in Oncology: ADEPT and the CRM," *Statistics in Medicine*, 27, 5345–5353.

Tozer, T. N. and Rowland, M. (2006), *Introduction to Pharmacokinetics and Pharmacodynamics: The Quantitative Basis of Drug Therapy*, Baltimore, MD: Lippincott Williams & Wilkins.