UNIVERSITY OF CALIFORNIA,
IRVINE

Bayesian Dose-Response Modeling in Sparse Data

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Steven B. Kim

Dissertation Committee:
Professor Daniel Gillen, Chair
Professor Wesley Johnson
Associate Professor Babak Shahbaba
Associate Professor Scott Bartell

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I appreciate all supports from my advisor, Dr. Gillen. He is not only an advisor in statistics, but he is also my advisor in my life. I also thank to the commitee members, Dr. Johson, Dr. Shahbaba and Dr. Bartell. I shall not forget special supports from my wife and parents.

I would like to thank for the support of NIA Grant T32-AG00096.

I thank Wiley & Sons for permission to include Chapter 4, which is in Risk Analysis, titled as Estimation of a Benchmark Dose in the Presence or Absence of Hormesis Using Posterior Averaging. The co-authors, Dr. Daniel Gillen and Dr. Scott Bartell, directed and supervised research which forms the basis for the dissertation.

CURRICULUM VITAE

Steven B. Kim

EDUCATION

Doctor of Philosophy in Statistics **2015**
University of California, Irvine *Irvine, California*

Master of Science in Applied Statistics **2010**
California State University, Long Beach *Long Beach, California*

Bachelor of Science in Mathematics **2007**
University of California, Irvine *Irvine, California*

RESEARCH EXPERIENCE

Trainee Researcher **2011–2014**
University of California, Irvine *Irvine, California*

Student Researcher **2010**
National Center for Toxicological Research *Jefferson, Arkansas*

TEACHING EXPERIENCE

Instructor **2015**
University of California, Irvine *Irvine, California*

Teaching Assistant **2010, 2014–2015**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Estimation of benchmark dose in the presence or absence of hormesis using posterior averaging **In press**
Risk Analysis

ABSTRACT OF THE DISSERTATION

Bayesian Dose-Response Modeling in Sparse Data

By

Steven B. Kim

Doctor of Philosophy in Statistics

University of California, Irvine, 2015

Professor Daniel Gillen, Chair

This book discusses Bayesian dose-response modeling in small samples applied to two different settings. The first setting is early phase clinical trials, and the second setting is toxicology studies in cancer risk assessment. In early phase clinical trials, experimental units are humans who are actual patients. Prior to a clinical trial, opinions from multiple subject area experts are generally more informative than the opinion of a single expert, but we may face a dilemma when they have disagreeing prior opinions. In this regard, we consider compromising the disagreement and compare two different approaches for making a decision. In addition to combining multiple opinions, we also address balancing two levels of ethics in early phase clinical trials. The first level is individual-level ethics which reflects the perspective of trial participants. The second level is population-level ethics which reflects the perspective of future patients. We extensively compare two existing statistical methods which focus on each perspective and propose a new method which balances the two conflicting perspectives. In toxicology studies, experimental units are living animals. Here we focus on a potential non-monotonic dose-response relationship which is known as hormesis. Briefly, hormesis is a phenomenon which can be characterized by a beneficial effect at low doses and a harmful effect at high doses. In cancer risk assessments, the estimation of a parameter, which is known as a benchmark dose, can be highly sensitive to a class of assumptions, monotonicity or hormesis. In this regard, we propose a robust approach which considers

both monotonicity and hormesis as a possibility. In addition, We discuss statistical hypothesis testing for hormesis and consider various experimental designs for detecting hormesis based on Bayesian decision theory. Past experiments have not been optimally designed for testing for hormesis, and some Bayesian optimal designs may not be optimal under a wrong parametric assumption. In this regard, we consider a robust experimental design which does not require any parametric assumption.

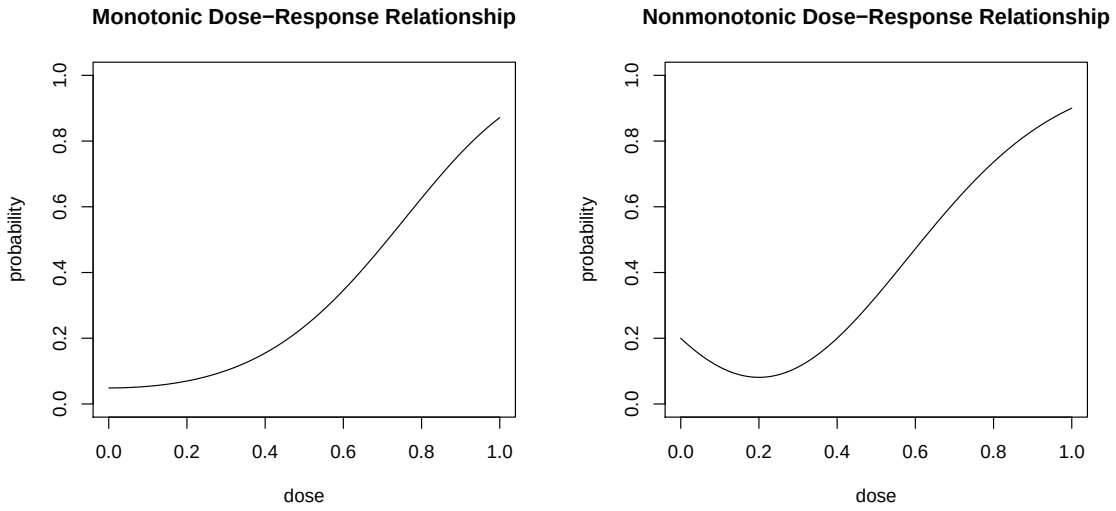
Chapter 1

Introduction and Background

The overarching goal of a dose-response model is to estimate the risk and/or benefit of a toxic agent or experimental treatment as a function of dose. Common situations where dose-response modeling is used include early phase clinical trials where interest often lies in estimating the maximal tolerable dose of a new investigational therapy, and toxicologic studies where the focus is on estimating tolerable levels of toxic agents found in the environment. In practice, a dose-response relationship may be monotonic, as shown in Figure 1.1a, or non-monotonic as shown in Figure 1.1b. Depending upon the scientific context, either assumption may be appropriate. Further, interest may lie anywhere from local approximation of a dose-response curve to classification of a dose-response type such as monotonicity or non-monotonicity.

The research presented here focuses on the commonly encountered scenario of estimating the dose-response relationship where outcomes are dichotomous. Examples include the occurrence of severe adverse events in the case of early phase clinical trials or the development of tumors in the case of environmental toxicology studies. Further, we focus on small-sample studies since most investigations concerned with dose-response assessment tend to rely on

Figure 1.1: Two classes of dose-response relationship



(a) Monotonic

(b) Nonmonotonic

relatively small sample sizes for logistical and/or ethical reasons. To overcome the inherent problems associated with statistical estimation in the presence of sparse data, the research presented throughout will primarily consider the development of Bayesian methods that allow for prior information elicited from prior research and investigator experience to be incorporated into the modeling process.

Among the various uses for dose-response modeling, our discussion mainly focuses on two previously mentioned topics: dose-response modeling in early phase clinical trials and toxicology studies for cancer risk assessment. Throughout, the overarching theme of the presented research is the development of flexible models for dose-response assessment that are robust to strong modeling assumptions. In clinical trials, experiments are performed on humans. In this case, we discuss methods that account for multiple prior opinions and the compromise between conflicting ethical perspectives. Since initiation of a clinical trial utilizing a Bayesian dose-finding design with a single investigator's prior "opinion" may lead to undesirable trial results, we consider gathering multiple prior opinions and accounting for prior uncertainty. In addition, we consider the development of statistical methods that explic-

itly address the potentially conflicting perspectives of individual- (i.e. trial patients) and population-level (i.e. future treated patients) ethics inherent in clinical trials design. The proposed research stresses the importance of an appropriate statistical strategy to balance these two perspectives.

In toxicology studies for cancer risk assessment, the experimental unit of interest is generally an animal. As such, efficient study design is essential for multiple ethical, economical and logistical reasons, while appropriate estimation techniques are critical to better ensure that meaningful conclusion may be drawn regarding potential results in human populations. With this in mind, we focus on novel research methods in three areas: (1) robust estimation of a safe dose for regulatory policy, (2) sensitive hypothesis testing procedures for determining monotonicity or non-monotonicity of a dose-response relationship, and (3) efficient experimental design methods for detecting a non-monotonic dose-response relationships when it is hypothesized that small levels of toxicologic agents may, in fact, be beneficial. Specifically, in toxicology and related fields, several scientists have argued that low exposure to a toxic substance may reduce the probability of a deleterious event when compared to no exposure to the substance, as depicted in Figure 1.1b. Despite this, most toxicology analyses either assume monotonicity of the dose-response curve or employ ad hoc procedures for determining non-monotonicity and subsequent dose-response estimation. The methods developed here address this gap in the literature by formally incorporating the possibility of non-monotonicity into both study design and estimation procedures.

The remainder of this chapter is dedicated to further familiarizing the reader with the background and setting of early phase clinical trials and toxicology studies for cancer risk assessment. While the primary contributions of this dissertation are to develop statistical methods, in this chapter we present background in the scientific applications that have motivated the work in order to understand and appreciate the perspectives presented.

1.1 Early Phase Clinical Trials

1.1.1 Introduction

In the US, a modern drug development process is divided into four phases. While later phases focus on efficacy of an experimental agent, the first phase (Phase I) focuses on safety of a new experimental drug when applied to human subjects. The primary objective of a Phase I clinical trial is to study toxicity of a new treatment and to determine an acceptable dose for later trials. A *maximum tolerable dose* (MTD) is defined as the highest dose that produces a desirable effect while avoiding unacceptable toxicity. A formal statistical definition of a MTD will be provided in Chapter 2 along with the basic notation used throughout Chapters 2 and 3. It is commonly assumed that both toxicity and efficacy increase with respect to dose. In this regard, neither underestimation nor overestimation of a MTD is desirable, and cost associated with mis-estimation in either direction depends heavily on the severity of the disease being treated. For example, in the context of developing new therapies for treating cancer, underestimation of a MTD decreases the likelihood of eliminating cancer cells, whereas overestimation of the MTD increases the likelihood of killing good cells.

As previously noted, most early phase clinical trials rely on relatively small sample sizes for logistical and ethical reasons. To overcome the inherent problems associated with statistical estimation in the presence of sparse data, it may be desirable to utilize Bayesian statistical methods that allow for prior information elicited from any research and investigator experience independent of current data. Multiple researchers have contributed to the development of Bayesian adaptive designs in Phase I clinical trials. However, Bayesian methods are still not widely used by trial investigators. While not fully known, the slow adoption of Bayesian methods in early phase clinical trials likely stems from the combination of a lack of understanding of Bayesian methods, computational complexity, model sensitivity, and prior sensitivity.

A focus on computational complexity is becoming less of a necessity since currently available computing power easily allows Bayesian adaptive designs and simulation studies. In addition, the potential negative impact from a misspecified parametric assumption often does not represent a great concern because the scientific goal in Phase I clinical trials is not to estimate the entire dose-response curve, but to estimate a local neighborhood of the MTD. In small-sample studies, we are able to primarily gain efficiency by borrowing from a mathematical framework posited by the dose-response model.

Prior sensitivity with respect to the distribution of the MTD and the distribution of the number of adverse events (AE) in a trial remains a real concern. One approach to avoiding the negative impact of prior sensitivity is to collect prior information from multiple experts instead of a single expert. To address prior sensitivity and to account for prior uncertainty, we develop and assess the method of weighting each posterior based on empirical evidence. We present theoretical justification and numerical experiments to illustrate the benefit of data-dependent weighting scheme in small-sample studies in Chapter 2.

While some argue that the ultimate goal of a clinical trial is to benefit future patients, trial participants are also real patients who should benefit from an experimental treatment as much as possible using as much knowledge as currently available regarding the safety and efficacy of the intervention. As such, we cannot ethically disregard the potential benefits and risks for trial participants during an experiment. We focus on two existing statistical design methods for Phase I clinical trials that have been previously proposed by O'Quigley et al. [45] and Whitehead and Brunier [67]. The two methods place differential weight on the ethical obligation of investigators to trial participants (so-call individual-level ethics) and to future treated patients (so-called population-level ethics). Specifically, the method proposed by O'Quigley et al. [45] tends to emphasize individual-level ethics, whereas the design procedure proposed by Whitehead and Brunier [67] tends to emphasize population-level ethics.

Palmer [48] extensively discussed the two different levels of ethics in clinical trials of all phases. However, the ethical issues inherent in clinical trial design and conduct are still, and will likely continue to be, debatable. If one makes a series of decisions to provide the best treatment to trial participants based on little accumulated knowledge in early stages of a clinical trial, information of the MTD will be slowly gathered. This result is intuitive because limited allocations over a restricted range of the dose space (the set of all possible experimental doses) limits knowledge regarding the overall dose-response curve. Hence, this strategy is not optimal for population-level ethics. On the other hand, if we make a series of decisions to maximize statistical information regarding the estimated MTD, under- and over-treatment of trial participants will frequently occur. Hence, the strategy of observing a wide range of dose space tends to allow population-level ethics to outweigh individual-level ethics. The novel methodology presented in Chapter 3 is motivated by our belief that there should exist a balance between the two levels of ethics in the statistical design and conduct of Phase I clinical trials and that such a balance should be formalized via statistical decision theory.

1.1.2 Further History on Statistical Methods for Phase I Clinical Trial Design

Markovian Up-and-Down Designs

Traditional designs for allocating a cohort of new patients to an experimental dose have historically been based on *up-and-down schemes*. Briefly, these designs are characterized by a dose allocating process that relies upon the outcomes of the last patient cohort treated at a given dose to determine whether the next patient cohort should move up in dose, move down in dose, or stay at the same dose. The most common of these is the 3 + 3 design (see Appendix A.1 for the dose-allocation algorithm). Such designs are generally known to be

conservative in that they tend to under-treat patients, yet are still widely used because they are mathematically and computationally simple to implement. Storer [60] used a Markov chain representation to evaluate conservative characteristics of multiple up-and-down rules. He showed that some simple alternatives to the traditional design [64, 66, 51] are just as conservative as the traditional 3 + 3 design. He showed that a two-stage design (transition from one alternative method to another alternative method) reduces bias in the estimation of a MTD.

Many statisticians have raised statistical and ethical concerns regarding the use of traditional up-and-down schemes. For example, a target MTD in the 3 + 3 design implicitly corresponds to a dose for which the probability of an AE is less than or equal to one-third. This definition of an acceptable toxicity is rather arbitrary and should clearly depend on the clinical setting being investigated. Further, the Markov property leads to both statistical and ethical concerns because it disregards a potentially large proportion of accumulated information in suggesting dose allocations for new patients. This concern has led to the proposal of several modified up-and-down schemes that include random-walk rules and accelerated titrations designs, though all still implicitly weight later observed observations more than earlier trial results [24, 28, 25, 59].

Bayesian Adaptive Designs

To overcome many of the statistical and ethical concerns of up-and-down schemes, O'Quigley et al. [45] proposed the *continual reassessment method* (CRM) for Phase I cancer studies. The CRM is a Bayesian adaptive design that places a prior distribution on the dose-response curve, in combination with all available data, when allocating the next patient to a new experimental dose. It was motivated by the underlying principle that trial participants should be treated at the current best guess of the most appropriate dose level based on all accumulated information. In this framework, a more precise definition of a MTD is provided

in a statistical sense. Given the more precise estimand, the CRM proceeds by treating a new patient at the posterior mean of the MTD, or at the closest experimental dose, therefore each patient in the trial has a relatively small chance of being allocated to a dose far away from the MTD when compared to the Markovian up-and-down designs.

The CRM immediately gained popularity in statistical and clinical communities, while some people worried about its relatively anti-conservative operating characteristics when compared to the traditional schemes. The general framework of the CRM has given rise to several proposed modifications in order to build in further safety constraints in dose escalations [27, 29, 3]. Goodman et al. [29] suggested limiting dose escalation by a single experimental dose level. Babb et al. [3] proposed to treat a new patient at a fixed quantile of the marginal posterior distribution of the MTD, and this method is known as the *escalation with overdose control* (EWOC). In the EWOC, a loss from over-dosing is greater than a loss from under-dosing of the same degree as opposed to the CRM which assigns the same loss for over-dosing and under-dosing of the same degree. The Frequentist approach of the CRM and the asymptotic property of the CRM were further investigated [46, 57], and some authors addressed prior sensitivity in the CRM [62, 70].

As previously noted, the CRM seeks to provide the best treatment to trial patients, which reflects individual-level ethics. Taking a different perspective than the CRM, Whitehead and Brunier [67] applied Bayesian decision theory to Phase I clinical trials by devising a dose-allocation scheme from the point of view of future patients. In this approach, the gain function was defined as the inverse of the asymptotic variance of the maximum likelihood estimator (MLE) for the MTD. This design is referred to as the *information gain method* (IGM) throughout our discussion. The IGM was devised from the perspective of population-level ethics as it seeks to optimize dose allocations to maximize precision of a resulting estimate of the MTD. The underlying idea is that future patients should be treated at a dose where the MTD is estimated as precisely as possible conditional upon the amount of

information one could reasonably obtain during a phase I clinical trial. This dose allocation rule requires a wide support of the dose space and tends to be applicable when the severity of an AE is mild to moderate (e.g. headache, rash, and easily reversible conditions) since dramatic increases and decreases in dose can be observed.

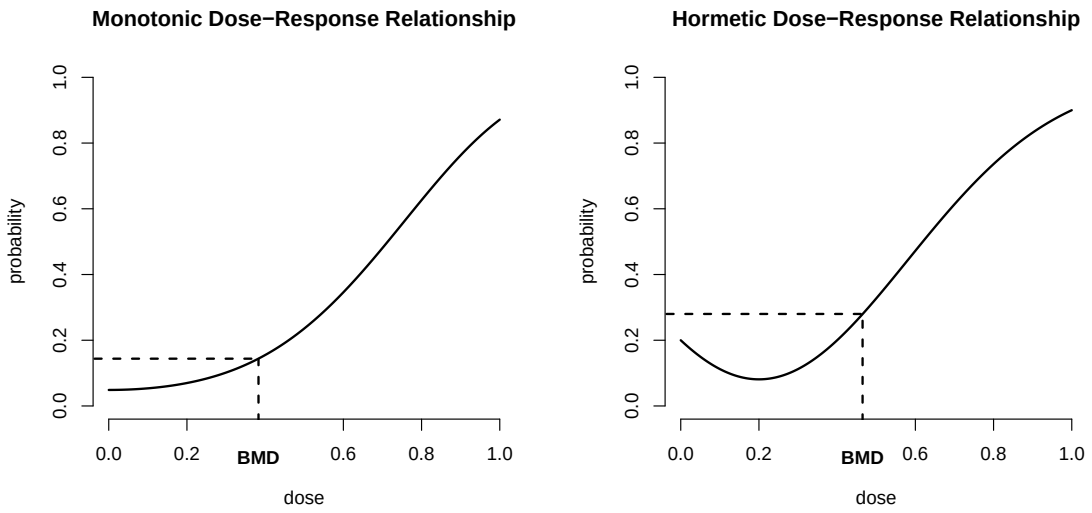
Whitehead and Williams [68] discussed more gain functions in the cases of cancer therapy and pointed out that conservatively modified CRMs may reduce the speed of the trial procedure. On the other hand, O'Quigley and Conaway [44] argued that, under current practice, it is not widely accepted to use a procedure that sacrifices the point of view of current trial participants to that of future patients. Following this, Bartroff and Lai [5] proposed a method that minimizes a loss function consisting of two additive terms, where each term reflects the individual and population risks. This approach attempted to compromise between the loss contributed by underdosing or overdosing a new patient and the incurred loss of the next patient who will receive a dose at the estimated loss function. This method reduced the complexity arising from considering infinitely many future patients by considering the next two patients.

1.2 Chemical Risk Assessments with Possibly Non-monotonic Dose-Response Relationships

1.2.1 Introduction

In toxicology, *hormesis* is defined as a dose-response relationship characterized by a low dose stimulation and a high dose inhibitory. It can be described by a J-shaped, U-shaped, or an inverted U-shaped dose-response curve. The theory of hormesis is based on the assumption of a stimulatory effect caused by low exposure to a toxic agent. Among various forms of

Figure 1.2: Benchmark doses under monotonicity and hormesis



(a) Monotonic

(b) Hormetic

hormesis, we particularly focus on a J-shaped hormetic curve which can be understood as a beneficial effect at low dose and a harmful effect at high dose, though beneficial and harmful effects are not necessary components of the definition of hormesis [13].

Some hormetic dose-response relationships are commonly known even outside toxicology. For example, adequate physical activity reduces oxidative stress whereas lack of and excessive physical activity increases the risk of high oxidative stress [49]. However, it is still controversial whether the theory of hormesis can be generalized to cancer-causing chemicals. Furthermore, a class of dose-response relationships depends on a specific endpoint of interest for a given toxic substance. For example, under the assumption of monotonicity with respect to toxicity and efficacy in chemotherapy, we can expect an inverted U-shaped dose-response relationship when the endpoint of interest is a conditional outcome such as a desirable event given no adverse event [69].

In cancer risk assessments, a *benchmark dose* (BMD) refers to the dose corresponding to a fixed relative increase in the risk of a deleterious event when compared to the background

risk (in the absence of an additional dose). The fixed relative increase is referred to as a *benchmark risk* (BMR) and is often fixed at 0.01, 0.05, or 0.10 in most assessments [26]. Figure 1.2 shows BMDs under monotonic and hormetic dose-response relationships when a BMR is fixed at 0.10. A statistical definition of a BMD given by the US Environmental Protection Agency (EPA) will be introduced in Chapter 4. An allowed risk in regulatory policy for public health is generally very small. However, estimation of such a low risk by an experiment requires an extremely large sample size. Though most cancer risk assessments are animal-based, such a large-sample study is considered to be ethically and logistically unacceptable. Because of this, it is common to first estimate a BMD for a reasonable BMR based on a small- or mid-sample study, then to use an estimated BMD as a point of departure for low dose extrapolation to protect public health. Monotonicity of the underlying dose-response curve has generally been assumed in this procedure. However, severe bias in the estimated BMD can result if the monotonicity assumption does not hold. To address this gap in methodology, the underlying methods presented in Chapter 4 allow for monotonic and hormetic dose-response relationships in the estimation process by extending the parameter space of commonly used multistage dose-response models in a novel way.

Hypothesis testing for hormesis has been of great scientific interest in toxicology. One difficulty in the testing arises from sparse experimental doses and sample sizes, particularly at low doses. It follows parametric testing procedures tend to suffer from high bias due to restricted, and often invalid, parametrization assumptions. To overcome these disadvantages, we consider a non-parametric approach using Bayesian model averaging [50]. We also consider a parametric approach which modifies the parameter space of a multistage dose-response model. The multistage model has long history in risk assessments, and it is one of several models equipped in the current EPA Benchmark Dose Software (BMDS). The modification is general so that it can be applied to other models in the BMDS. A detailed investigation is presented in Chapter 5.

Many past experiments were not optimally designed to assess a hormetic effect at low dose. For future data collection to test hormesis, it is important to consider optimal dose allocations. Zapponi and Marcello [71] reviewed multiple datasets that exhibit potential evidence for hormesis in cancer-causing agents, but a common experimental design in the presented datasets is based on nearly equal sample sizes at given experimental doses. If we allocate 180 animals to six experimental doses, say (0.0, 0.2, 0.4, 0.6, 0.8, 1.0), we expect different degrees of evidence for hormesis from the study design of (50, 50, 50, 10, 10, 10) and the study design of (30, 30, 30, 30, 30, 30). As such, we discuss the importance of an experimental design in Chapter 6 and propose various experimental designs based upon Bayesian decision theoretic approaches in order to maximize statistical information for quantifying the presence of a hormetic dose-response relationship.

1.2.2 Further History on Statistical Methods for Toxicology Studies of Carcinogen Risk

BMD Estimation and Regulatory Policy

For protecting public health, accurate estimation of a BMD is important as an estimated BMD serves as a point of departure for further downward extrapolation. One difficulty in estimation of a BMD for a low BMR arises from sparseness of data particularly at low dose levels. A regression technique allows us to interpolate between two experimental doses or to extrapolate from the lowest experimental dose, but the resulting estimate of the BMD is highly sensitive to model specification even within a class of monotonic dose-response models. This follows because a regressed dose-response curve behaves uniquely at low dose from model to model. For certain monotonic dose-response models derived based on biological interpretations, such as a one-parameter exponential model and a two-parameter Beta-Poisson model, it can be shown that a dose-response relationship at arbitrarily low dose is linear [30].

Therefore, a linear downward extrapolation from an estimated BMD has been practiced for regulatory purposes.

To address model uncertainty, implementation of frequentist model averaging techniques has been extensively discussed in the risk assessment based community. Weights for model averaging include various information criteria including the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Kullback information criterion [9, 65, 4, 41]. A common framework based on these information criteria increases the contribution of a well fitted model and decreases the contribution of a complex model. Recently, implementation of BMA in risk assessments has been discussed [56]. In the aforementioned studies, however, only monotonic dose-response models are considered for model averaging.

Allowing of Hormesis in BMD Estimation

In contrast to conservative estimation approaches based on monotonicity that have been taken by several regulatory agencies, some scientists have argued the existence of hormesis and have even raised the importance of accepting hormesis as a default dose-response relationship in risk assessments. Calabrese and Baldwin [11] quantified evidence of hormesis and categorized approximately 350 studies into high, moderate, and low hormetic effects. Following this work, multiple studies have elevated the importance of hormesis in toxicology by reporting empirical evidence for hormesis using a series of ad hoc methodologies [12, 14, 15, 19]. In particular, Calabrese and Cook [15] discussed various advantages to parameter estimation if the possibility of a hormetic function was accepted as the default assumption in carcinogen risk assessments. One possible advantage they pointed out is the protection of both normal and high-risk populations by estimating a BMD where the corresponding risk is lower than the background risk for both groups. Further, Zapponi and Marcello [71] reviewed several datasets which was appealed to be evidence of hormesis.

Despite the long history of hormesis theory, it has not been widely accepted by regulatory agencies such as the EPA. Despite valid points regarding the costs of ignoring the possibility of hormesis, some authors have pointed out limitations of the ad hoc estimation approaches and the lack of validated hypothesis testing procedures applied to existing database [61, 42]. It has been argued that the current practice of using a lower confidence limit of a BMD, referred to as BMDL, or an averaged BMDL from various monotonic models is conservative and health-protective, particularly when evidence for hormesis is not definitive [61, 26].

A multistage model, which is the foundation for the development of the following statistical methods, was derived from biological mechanisms of action. Armitage and Doll [2] discussed the two-stage theory of carcinogenesis, and Armitage [1] argued that a multistage model with several stages (e.g. five to seven stages) has often been regarded as implausible in the absence of direct biological evidence about a succession of stages. He further mentioned that one can either assume more than two stages or regard several intermediate stages as a single phase as approximation. A biological interpretation is important, but a model with such a large number of parameters may suffer from over-fitting in sparse data and experimental doses. Crump [20] reviewed a linearized multistage model, and Kuo and Cohen [39] presented a Gibbs algorithm for the multistage model under the monotonic assumption.

There are many flexible dose-response models which are suitable for modeling hormesis in dichotomous and continuous response [40, 32, 33, 53, 7, 8]. Many of these flexible models are not currently installed in the EPA BMDS.

Chapter 2

An Alternative Perspective on the Use of Consensus Estimators

In Chapter 2, we focus on the incorporation of multiple priors in Bayesian adaptive studies. We consider two classes of Bayes estimators. The first class averages individual Bayes estimators with the weight initially determined before data collection. In the second class of estimators, a single consensus prior is first formed as a linear combination of weighted priors, and the resulting Bayes estimator is then obtained. We show that this class results in a natural posterior weighting scheme with prior weights updated conditional upon observed data. We focus on the case when we periodically analyze data and sequentially make decisions based on observed data.

2.1 Introduction

In small-sample studies, statisticians often rely on parametric assumptions to gain efficiency for estimation and testing. In a Bayesian framework, additional prior information is also in-

corporated into the model and statistical inference regarding parameters of scientific interest is based on the resulting posterior distribution, which is obtained by updating the prior with observed empirical data. In many cases the posterior distribution of a parameter of interest will depend on the prior elicitation in an appreciable way, particularly in a sparse data setting. It can be particularly critical in the setting of Phase I clinical trials because both dose allocations for trial participants and an estimated dose appropriate for future patients may significantly depends on the prior elicitation.

When eliciting prior information, knowledge from multiple subject area experts are generally more informative than the knowledge of a single expert. On the other hand, when multiple experts have divergent prior information, we face the dilemma of how to combine this information into inferential procedures. This issue is most important in sparse data settings where inference may change substantially depending upon the way the information is incorporated into the model. The incorporation of multiple prior opinions has previously been considered and many advocate that simple combination methods tend to perform reasonably well relative to more complicated approaches [18]. One approach for incorporating multiple prior opinions focuses on the use of *consensus estimators*, as discussed by Samaniego [52], among others. Samaniego focused on a consensus class of estimators that combines individual posterior estimates [52]. When each individual has a unique posterior estimate originating from a unique prior distribution, this class of estimators averages the multiple posterior estimates based on the credibility of each prior elicitation determined before observing any data. This approach would seem more robust than relying on a single posterior estimate as it marginalizes uncertainty in the prior across various experts.

When it is fairly evident that one expert's prior information is far more plausible, after observing data, relative to the other prior information, it is tempting to determine the credibility of each prior based on the observed data. In many cases, such an approach would result in foundationally unsound inference as it would constitute "double dipping" of

the data by modifying the pre-specified prior distribution with information obtained during the experiment. However, in this chapter we provide an alternative perspective of a single pre-specified consensus prior that is formed as a, possibly weighted, linear combination of the available priors. Specifically, we show that the use of the consensus prior results in a natural posterior weighting scheme with individual weights based on observed data. The result is particularly useful in the setting of adaptive studies where design modifications are periodically made as new data are observed, since it demonstrates that future design decisions tend to shrink towards the most plausible prior when the consensus prior is used.

In Section 2.2, we present two classes of consensus estimators, while briefly reviewing a consensus prior and the notion of self-consistency in the context of Bayesian estimation [52]. We then provide an alternative perspective of posterior estimators based on the consensus prior, showing that they are equivalent to a class of estimators that weight the contribution of each prior conditionally upon all observed data. We also comment on Occam's window criterion in the context of a consensus prior. In Section 2.3, we provide simple examples in the context of a normal probability model to enhance the understanding of the alternative perspective of a consensus prior. In Section 2.4, we consider the use of consensus estimators in the context of early phase clinical trials that necessitate adaptive dose allocations. Simulation studies are presented to illustrate the utility of incorporating multiple prior opinions and to compare the operating characteristics of consensus estimators in the adaptive phase I trial setting. We conclude with a discussion on the utility of consensus priors in scientific investigations in Section 2.5.

2.2 Bayesian Decisions Based on a Consensus Prior

2.2.1 A Consensus Prior

In small-sample studies a single prior that is contrary to observed data may substantially influence posterior inference, and hence scientific decisions based on the resulting posterior distribution. To guard against this undesirable situation, we may rely on multiple opinions. In this setting it is necessary to incorporate prior disagreements in the resulting inferential procedure.

We initially consider a case where all investigators specify conjugate priors through the same model (i.e. the same likelihood function). Let $\mathcal{Q} = \{Q_1, \ldots, Q_K\}$ denote a set of priors and $g(\theta \mid Q_k)$ denote the prior density function of θ from Q_k. As discussed by Samaniego [52], one possible approach for a compromise is to gather the multiple investigators and formulate a single prior through conversations. However, this process may not be systematic and is logistically infeasible in many settings. A simple alternative and systematic approach to account for multiple priors is to specify a K-fold mixture

$$g_*(\theta) = \sum_{k=1}^{K} g(\theta \mid Q_k)\, P(Q_k)\,, \tag{2.1}$$

which is referred to throughout as a *consensus prior*. The pre-assigned prior probabilities $P(Q_k)$, $Q_k \in \mathcal{Q}$, reflect the plausibility of each investigator's prior, perhaps depending on various degrees of knowledge and experience. Multiple advantages stemming from the use of a consensus prior based on individual conjugate priors are listed in Samaniego [52]. One particular advantage of interest is that the posterior distribution is also a K-fold mixture of the conditional posteriors $g(\theta \mid \vec{y}, Q_k)$, where $\vec{y}$ denotes observed data. Its mathematical properties will be further discussed in Section 2.3.

2.2.2 Self-Consistency

Suppose that $\hat{\theta}$ is a sufficient statistic and is an unbiased estimator of a scalar parameter θ. The Bayes estimator of θ that minimizes squared error loss is said to be *self-consistent* if $E(\theta \mid \hat{\theta} = \theta^*) = \theta^*$, where $\theta^* = E(\theta)$ is the prior expectation [52]. If such an estimator exists, the property of self-consistency is appealing in the sense that the posterior estimate is θ^* when both the prior estimate and the unbiased estimator are θ^*. While self-consistency is not guaranteed in general, particularly when the estimand of interest is a nonlinear function of regression parameters, we can find a self-consistent estimator of the form $(1-\eta)\theta^* + \eta\hat{\theta}$ under some simple models.

Using the form of a consensus prior in Equation (2.1), Samaniego [52] noted that self-consistency is not preserved in general. He instead discussed the class of estimators of the form

$$\hat{\theta}^* = \sum_{k=1}^{K} w_k \hat{\theta}_k \,, \tag{2.2}$$

where an individual posterior estimator $\hat{\theta}_k$ is of the form $(1-\eta_k)\theta_k^* + \eta_k\hat{\theta}$. We refer to this class of estimators as *weighted posterior estimators*. A weighted posterior estimator is a natural compromise when individual posterior estimates disagree, where the contribution of each investigator is determined by its prior plausibility, say $w_k = P(Q_k)$. If η_k is constant in ks, the property of self-consistency is preserved because

$$E\left(\theta \mid \hat{\theta} = \theta^*\right) = (1-\eta)\sum_{k=1}^{K} w_k\,\theta_k^* + \eta\sum_{k=1}^{K} w_k\,\theta^* = (1-\eta)\,\theta^* + \eta\,\theta^* \,. \tag{2.3}$$

However, if η_k depends on k, self-consistency is not generally guaranteed.

2.2.3 An Alternative Perspective on the Consensus Prior

While the notion of self-consistency is mathematically and intuitively appealing to some, there are cases when it is tempting to use empirical evidence to determine each individual's contribution to posterior estimation. For instance, when $\hat{\theta}_1 = -1$ and $\hat{\theta}_2 = 1$ after observing $\hat{\theta} = 2$, it may be a natural reaction to desire $w_2 > w_1$. Data-dependent weighting schemes are no longer Bayesian, but rather are Empirical Bayesian, since the data are used in the construction of the prior. A purely Bayesian approach is rigorous that prior information should be obtained independent of the data. While briefly noted in Samaniego [52], in this section we explicitly show that at least one data-dependent weighting scheme can be formally justified by showing its equivalence to the use of a consensus prior that is specified prior to the observation of any experimental data. The result not only provides formal justification for the use of adaptively weighted priors in sequential experimentation but also yields further intuitive appeal for the use of consensus priors in Bayesian inference.

Let $f(\vec{y} \mid \theta)$ be the likelihood function, which is the same regardless $Q_k \in \mathcal{Q}$. The marginal likelihood function

$$f(\vec{y} \mid Q_k) = \int f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)\, d\theta \tag{2.4}$$

does depend on Q_k. Given $\vec{y}$, we can quantify the updated belief for $Q_k \in \mathcal{Q}$ by

$$P(Q_k \mid \vec{y}) = \frac{f(\vec{y} \mid Q_k)\, P(Q_k)}{\sum_{j=1}^{K} f_j(\vec{y} \mid Q_j)\, P(Q_j)}\,. \tag{2.5}$$

This unconventional quantity thus represents the posterior probability of prior Q_k. Conditioning on Q_k, the individual posterior probability distribution of θ is

$$g(\theta \mid \vec{y}, Q_k) = \frac{f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)}{f(\vec{y} \mid Q_k)}\,. \tag{2.6}$$

In the following proposition we claim that the weighted average of the individual posterior probability density functions

$$g_2(\theta \mid \vec{y}) = \sum_{k=1}^{K} g(\theta \mid \vec{y}, Q_k)\, P(Q_k \mid \vec{y}) \tag{2.7}$$

is legitimate in the sense that it is equivalent to

$$g_1(\theta \mid \vec{y}) = \frac{f(\vec{y} \mid \theta)\, g_*(\theta)}{\int f(\vec{y} \mid \theta)\, g_*(\theta)\, d\theta}\,, \tag{2.8}$$

where $g_*(\theta)$ is the consensus prior in Equation (2.1).

Proposition. For a given set of priors $g(\theta \mid Q_k)$ for $k = 1, \ldots, K$, we have $g_1(\theta \mid \vec{y}) = g_2(\theta \mid \vec{y})$.

Proof. By Equations (2.1) to (2.7),

$$\begin{aligned}
g_2(\theta \mid \vec{y}) &= \sum_{k=1}^{K} g(\theta \mid \vec{y}, Q_k)\, P(Q_k \mid \vec{y}) \\
&= \sum_{k=1}^{K} \frac{f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)}{f(\vec{y} \mid Q_k)}\, \frac{f(\vec{y} \mid Q_k)\, P(Q_k)}{\sum_{j=1}^{K} f_j(\vec{y} \mid Q_j)\, P(Q_j)} \\
&= \sum_{k=1}^{K} \frac{f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)\, P(Q_k)}{\sum_{j=1}^{K} f_j(\vec{y} \mid Q_j)\, P(Q_j)} \\
&= \sum_{k=1}^{K} \frac{f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)\, P(Q_k)}{\sum_{j=1}^{K} \left(\int f(\vec{y} \mid \theta)\, g_j(\theta \mid Q_j)\, d\theta\right)\, P(Q_j)} \\
&= \frac{f(\vec{y} \mid \theta)\, \sum_{k=1}^{K} g(\theta \mid Q_k)\, P(Q_k)}{\int f(\vec{y} \mid \theta)\, \left(\sum_{j=1}^{K} g_j(\theta \mid Q_j)\, P(Q_j)\right)\, d\theta} \\
&= \frac{f(\vec{y} \mid \theta)\, g_*(\theta)}{\int f(\vec{y} \mid \theta)\, g_*(\theta)\, d\theta} \\
&= g_1(\theta \mid \vec{y})\,.
\end{aligned}$$

Therefore, posterior inference obtained by using the pre-specified consensus prior is equivalent to the posterior inference based on the weighted posterior density function. In addition, the posterior contribution of the k^{th} investigator is exactly equal to $P(Q_k \mid \vec{y}_n)$ in Equation (2.5). Equation (2.7) is reminiscent of Bayesian model averaging, which involves a similar posterior weighted average [50].

Based upon the weighted posterior density function, $g_2(\theta \mid \vec{y})$, the resulting Bayes estimator of θ which minimizes squared error loss is

$$\hat{\theta}^{**} = \sum_{k=1}^{K} w_k^* \hat{\theta}_k \,,$$

where $w_k^* = P(Q_k \mid \vec{y})$ is the posterior weight given in Equation (2.5) and $\hat{\theta}_k = E(\theta \mid \vec{y}, Q_k)$ is the individual Bayes estimator that minimizes the individual posterior squared error loss. Law of Total Probability establishes validity of the estimator since it is equivalent to the Bayes estimator obtained from use of the pre-specified consensus prior. As such, we refer to this estimator as the *consensus prior estimator*.

2.2.4 A Note on the Occam's Window Criterion

In this section, we address the importance of maintaining all prior specifications until the end of Bayesian inference. Though the weighting methods in $\hat{\theta}^*$ and $\hat{\theta}^{**}$ are different, both methods preserve an assigned prior probability distribution on $\mathcal{Q}$. Some authors have proposed the use of Occam's window criterion to refine the model space in the context of Bayesian model averaging when a posterior decision is required [70]. In this case, it is argued that the elimination of poorly fitted model(s) from the model space (i.e. set of candidate models) is reasonable if the fit of a model is far worse than the fit of the best-fitting model. In the context of a consensus prior, it may be also tempting to take this approach by eliminating

$Q_k \in \mathcal{Q}$ if

$$\frac{P(Q_j \mid \vec{y})}{\max_j P(Q_j \mid \vec{y})} < \epsilon \tag{2.9}$$

for some prefixed $\epsilon > 0$ according to the Occam's window criterion. The impact of such a poor prior is down-weighed through the consensus prior (and through the consensus posterior), but the elimination of any $Q_k \in \mathcal{Q}$ based on the observed data would lead to changing the prior probabilities assigned to the initial set $\mathcal{Q}$. To see this, we let $E_k(\vec{y})$ denote the complementary of the event described in Equation (2.9) (i.e. Q_k is not eliminated given $\vec{y}$). Attempting to refine $\mathcal{Q}$, the posterior density function of θ is given by

$$g(\theta \mid \vec{y}) = \sum_{k=1}^{K} g(\theta \mid \vec{y}, Q_k)\, P^*(Q_k \mid \vec{y}) \tag{2.10}$$

with the normalized posterior weight

$$P^*(Q_k \mid \vec{y}) = \frac{P(Q_k \mid \vec{y})\, \mathrm{I}_{E_k(\vec{y})}}{\sum_{j=1}^{K} P(Q_j \mid \vec{y})\, \mathrm{I}_{E_j(\vec{y})}}\,, \tag{2.11}$$

where $\mathrm{I}_E = 1$ if an event E occurs, zero otherwise.

Proposition. The application of the Occam's window criterion to a consensus prior is not coherent.

Proof. From Equations (2.5) and (2.9) to (2.11),

$$
\begin{aligned}
g(\theta \mid \vec{y}) &= \sum_{k=1}^{K} \frac{f(\vec{y} \mid \theta, Q_k)\, g(\theta \mid Q_k)}{f(\vec{y} \mid Q_k)} \frac{P(Q_k \mid \vec{y})\, \mathrm{I}_{E_k(\vec{y})}}{\sum_{j=1}^{K} P(Q_j \mid \vec{y})\, \mathrm{I}_{E_j(\vec{y})}} \\
&= \sum_{k=1}^{K} \frac{f(\vec{y} \mid \theta)\, g(\theta \mid Q_k)}{f(\vec{y} \mid Q_k)} \frac{\frac{f(\vec{y}|Q_k)\, P(Q_k)}{\sum_{j=1}^{K} f(\vec{y}|Q_j)\, P(Q_j)}\, \mathrm{I}_{E_k(\vec{y})}}{\sum_{j=1}^{K} \frac{f(\vec{y}|Q_j)\, P(Q_j)}{\sum_{i=1}^{K} f(\vec{y}|Q_i)\, P(Q_i)} \mathrm{I}_{E_j(\vec{y})}} \\
&= f(\vec{y} \mid \theta) \sum_{k=1}^{K} \frac{g(\theta \mid Q_k)\, P(Q_k)\, \mathrm{I}_{E_k(\vec{y})}}{\sum_{j=1}^{K} f(\vec{y} \mid Q_j)\, P(Q_j)\, \mathrm{I}_{E_j(\vec{y})}} \\
&= \frac{f(\vec{y} \mid \theta) \sum_{k=1}^{K} g(\theta \mid Q_k)\, P(Q_k)\, \mathrm{I}_{E_k(\vec{y})}}{\sum_{j=1}^{K} \left(\int f(\vec{y} \mid \theta)\, g(\theta \mid Q_j)\, d\theta \right) P(Q_j)\, \mathrm{I}_{E_j(\vec{y})}} \\
&= \frac{f(\vec{y} \mid \theta) \sum_{k=1}^{K} g(\theta \mid Q_k)\, P(Q_k)\, \mathrm{I}_{E_k(\vec{y})}}{\int f(\vec{y} \mid \theta) \sum_{j=1}^{K} g(\theta \mid Q_j)\, P(Q_j)\, \mathrm{I}_{E_j(\vec{y})}\, d\theta} .
\end{aligned}
$$

Therefore, the data-dependent weighted prior density

$$
\sum_{k=1}^{K} g(\theta \mid Q_k)\, P(Q_k)\, \mathrm{I}_{E_k(\vec{y})}
$$

cannot be considered as a consensus prior because it depends on $\vec{y}$.

Similarly, it is worth noting that the same concern arises when one applies the Occam's window criterion in the context of Bayesian model averaging. In addition, the application precludes self-consistency of $\hat{\theta}^*$ in Equation (2.2) with constant $\eta_k = \eta$ because the last equality in Equation (2.3) does not hold with altered w_k. From the perspective of self consistency of $\hat{\theta}^*$ and the perspective of the Bayes estimator $\hat{\theta}^{**}$, all priors should be incorporated into the analysis.

2.3 Examples

2.3.1 Normal Model with Unknown Mean and Known Variance

We consider a random sample $\vec{Y} = (Y_1, \ldots, Y_n)$, where $Y_i \sim N(\mu, \sigma^2)$ with unknown μ and known σ^2. Under this exponential family, the sample mean $\hat{\mu} = \bar{Y}$ is sufficient and unbiased for μ. For $k = 1, \ldots, K$, the conjugate priors $\mu \mid Q_k \sim N(\mu_k^*, \sigma_k^2)$ leads to the posterior $\mu \mid \vec{y}, Q_k \sim N\left(\hat{\mu}_k, \tau_k^{-1}\right)$, where $\tau_k = \sigma_k^{-2} + n\sigma^{-2}$ denotes the posterior precision (i.e. inverse of the posterior variance) and the Bayes estimator $\hat{\mu}_k$ is in the form of $(1 - \eta_k)\mu_k^* + \eta_k \bar{Y}$ with $\eta_k = n\sigma^{-2}\tau_k^{-1}$. For self-consistency, we require the same prior precision $\tau_k = \tau$, and the self-consistent weighted posterior estimator of μ is

$$\hat{\mu}^* = \sum_{k=1}^{K} w_k \, \hat{\mu}_k \,,$$

where $w_k = P(Q_k)$.

In general, each individual marginal likelihood function can be written as

$$f(\vec{y} \mid Q_k) \propto \frac{1}{\sqrt{v_k}} \exp\left(\frac{(\bar{y} - \mu_k^*)^2}{2v_k}\right) ,$$

where $v_k = \sigma_k^2 + \sigma^2 n^{-1}$. To this end, the posterior contribution of Q_k to the weighted posterior $g_2(\mu \mid \vec{y})$ is

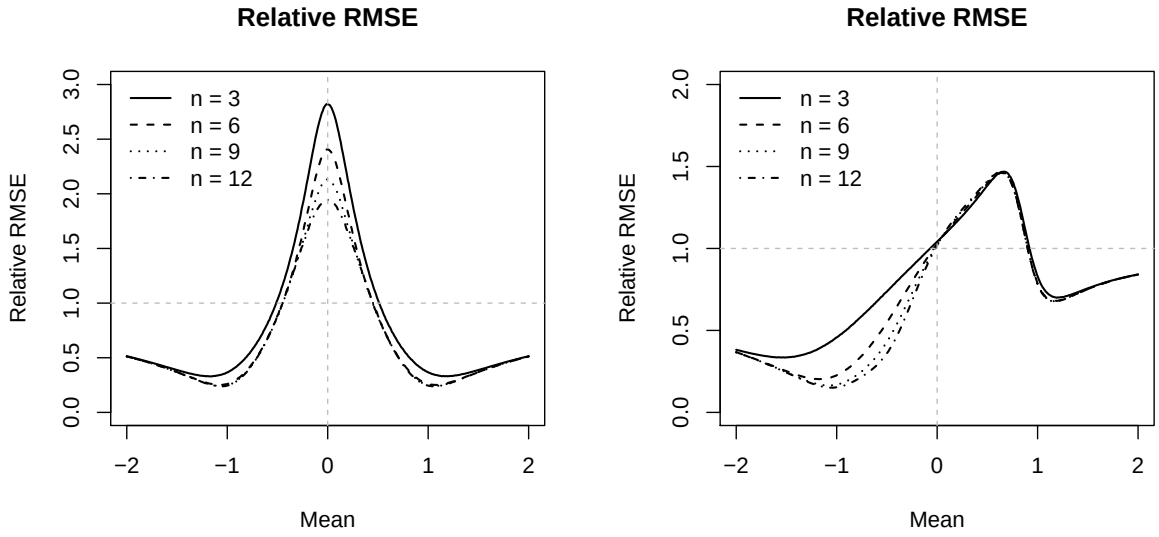
$$P(Q_k \mid \vec{y}) \propto P(Q_k) \, \phi(\bar{y};\, \mu_k^*, v_k) \,,$$

where $\phi(\cdot)$ denotes a normal density function. Then, the consensus prior estimator of θ is

$$\hat{\mu}^{**} = \sum_{k=1}^{K} w_k^* \, \hat{\mu}_k \,,$$

with $w_k^* = P(Q_k)\,\phi(\bar{y};\,\mu_k^*, v_k)$. In practice, one expert may be more experienced and hence may receive a relatively greater prior weight $P(Q_k)$. However, for a given $P(Q_k)$, a strong prior expressed via a small value of σ_k^2 additionally increases the posterior weight w_k^*. This is a typical behavior in the posterior weighting scheme. As such, one might consider whether to upweight an individual opinion simply because of a strong prior.

Figure 2.1: Mean squared error with respect to the mean under the normal model



(a) $P(Q_1) = P(Q_2)$ (b) $P(Q_2) = 9 \times P(Q_1)$

Typically with a small sample, the bias of $\hat{\mu}^{**}$ is expected to be small relative to the bias of $\hat{\mu}^*$ because $\hat{\mu}^{**}$ incorporates weights that have been adjusted based on observed data. On the other hand, the variance of $\hat{\mu}^{**}$ is large relative to the variance of $\hat{\mu}^*$ because the data-dependent weight w_k^* is random. As a numerical example for the trade-off between bias and variance, we consider $K = 2$ priors with $\sigma_k = 0.5$ (i.e. prior sample size of two) and $\mu_k^* = (-1)^k$ for $k = 1, 2$. Figure 2.1 shows the curve of relative root mean squared error (RMSE) of $\hat{\mu}^{**}$ to $\hat{\mu}^*$ when μ varies from -2 to 2 for small sample sizes $n = 3, 6, 9, 12$. Focusing on the case of $P(Q_1) = P(Q_2)$ (see Figure 2.1a), RMSE exceeds one (i.e. the self-consistent weighted posterior estimator, $\hat{\mu}^*$, performs better) when μ is close to the midpoint of the two prior estimates. On the other hand, RMSE is below one (i.e. the consensus prior estimator

$\hat{\mu}^{**}$ performs better) when μ is substantially closer to one of the two prior estimates. The relative performance of $\hat{\mu}^{**}$ increases under all values of μ as n increases. Focusing on the case of $P(Q_2) = 9 \times P(Q_1)$ (see Figure 2.1b), a similar trend is shown with the zone of superiority of $\hat{\mu}^{*}$ shifted to $\mu_2^* = 1$. When the true value of μ is near μ_2^*, the consensus prior estimator $\hat{\mu}^{**}$ performs better with respect to RMSE. When the true value of μ is against the prior assignment $P(Q_2) = 9 \times P(Q_1)$ (i.e. μ is close to μ_1^*), RMSE decreases, and the rate of decrease is positively correlated with the sample size.

2.3.2 Normal Model with Unknown Mean and Variance

We now assume both μ and σ^2 are unknown and consider the conjugate normal-gamma prior

$$\mu \mid \tau, Q_k \sim N\left(\mu_k^*, \frac{1}{n_k \tau}\right) \,, \;\; \tau \mid Q_k \sim \text{Gamma}(a_k, b_k) \,,$$

where $\tau = \sigma^{-2}$. The parameter of interest is μ, and σ^2 is the nuisance parameter. It can be shown that the k^{th} marginal posterior distribution of μ is given by

$$\mu \mid \vec{y}, Q_k \sim T\left(2a_k^*, \hat{\mu}_k, \frac{a_k^* \left(n_k + n\right)}{b_k^*}\right)$$

with

$$a_k^* = a_k + \frac{n}{2} \,, \;\; b_k^* = b_k + \frac{1}{2}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{n_k n \left(\bar{y} - \mu_k^*\right)^2}{n_k + n}\right) \,, \;\; \hat{\mu}_k = \frac{n_k \mu_k^* + n\bar{y}}{n_k + n} \,.$$

Therefore, a self-consistent weighted posterior estimator, $\hat{\mu}^*$, can be achieved by equating n_k.

The k^{th} marginal likelihood is given by

$$f(\vec{y} \mid Q_k) = \frac{\Gamma(a_k^*)}{\Gamma(a_k)} \frac{(b_k)^{a_k}}{(b_k^*)^{a_k^*}} \left(\frac{n_k}{n_k + n}\right)^{1/2} \left(\frac{1}{2\pi}\right)^{n/2} .$$

To simplify the computation of the consensus prior estimator, $\hat{\mu}^{**}$, we may consider constant (a_k, b_k, n_k) while varying $P(Q_k)$ if necessary. In this case, the difference among priors is expressed only through μ_k^*, and the posterior weight becomes

$$P(Q_k \mid \vec{y}_n) \propto P(Q_k) \frac{(b_k^*)^{-a_k^*}}{\sum_{j=1}^{K} (b_j^*)^{-a_j^*}} ,$$

where $(\bar{y} - \mu_k^*)^2$ is an important determinant in b_k^*. Thus the closer the distance between the summary statistic and the prior guess the greater the contribution to the posterior. As such, the relative operating characteristics of the posterior weighted estimator, $\hat{\mu}^*$, and the consensus prior estimator, $\hat{\mu}^{**}$, are similar to the case of known variance.

2.4 Incorporation in Adaptive Phase I Clinical Trials

A maximum tolerable dose (MTD) is the highest dose of a therapeutic treatment that does not cause unacceptable toxicity in a loose definition. A precise definition will be provided later in this section. A primary objective of most Phase I clinical trials is to study the toxicity of a new drug and to determine the MTD for later investigation and future patients. Whitehead and Williamson [68] discussed various Bayesian decision theoretic approaches for dose-finding studies based on a logistic regression model. Among various gain functions discussed, we focus on the *patient gain* using the terminology in the paper. For the patient gain, the dose allocation rule is optimal for each member of trial participants. This experimental design is analogous to the continual reassessment method (CRM) proposed by O'Quigley et al. [45], which was first developed to treat cancer patients in severe conditions. One major

concern about this Bayesian adaptive design is prior sensitivity. In particular, estimation of the MTD and the number of adverse events (AEs) per trial can be highly sensitive to a prior specification because we often have thirty or fewer subjects in Phase I clinical trials. Sometimes, multiple researchers may have different opinions about the toxicity of a new experimental drug while the dose allocations for the first few trial participants heavily depend on their subjective information. Therefore, the application of a consensus prior may be recommended.

2.4.1 Elicitation and Incorporation of Multiple Prior Opinions

Suppose we observe $\vec{Y}_n = (Y_1, \ldots, Y_n)$, where $Y_i \sim \text{Bernoulli}(\pi_{x_i})$ with $Y_i = 1$ indicating an AE at the treated dose x_i in log-scale. Using a logistic regression, the likelihood function of $\vec{\beta} = (\beta_0, \beta_1)$ is

$$f(\vec{y} \mid \vec{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{1-y_i}$$

by assuming independence among patients. We let $\beta_0 \in (-\infty, \infty)$ and $\beta_1 \in (0, \infty)$ by assuming a monotonic dose-response relationship. We consider a conditional mean prior [67, 6]. This method of prior elicitation requires the selection of two arbitrarily doses, say $x_{-1} < x_0$ without loss of generosity. Then, we specify two independent Beta priors $\pi_{x_i} \sim \text{Beta}(a_i, b_i)$ for $i = -1, 0$. By the Jacobian transformation, we obtain the joint prior density function

$$f(\vec{\beta}) = \prod_{i=-1}^{0} \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\,\Gamma(b_i)} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{a_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{b_i} (x_0 - x_{-1}),$$

hence a_i and b_i can be thought of as the pseudo numbers of AEs and non-AEs, respectively, at the selected dose x_i. To this end,

$$f(\vec{\beta} \mid \vec{y}) \propto \prod_{i=-1}^{n} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{1-y_i} (x_0 - x_{-1}),$$

where $y_i = a_i$ and $1 - y_i = b_i$ for $i = -1, 0$.

If we have multiple experts we may ask each k^{th} investigator independently to specify (x_{ik}, a_{ik}, b_{ik}) for $i = -1, 0$, thus defining the set of priors, $\mathcal{Q} = \{Q_1, \ldots, Q_K\}$. In addition to conjugacy, the conditional means prior is advantageous because it is more interpretable than a direct specification of the joint prior density function of $\vec{\beta}$. Hence, it is easy to control the amount of prior information in a consensus prior. The weighted posterior density function is

$$g(\vec{\beta} \mid \vec{y}) \propto \sum_{k=1}^{K} P(Q_k \mid \vec{y}) \prod_{i=-1}^{n} c_{ik} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_{ik}} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{n_{ik}-y_{ik}},$$

where $y_{ik} = y_i$, $n_{ik} = 1$, and $c_{ik} = 1$ for $i = 1, \ldots, n$, and $y_{ik} = a_{ik}$, $n_{ik} = a_{ik} + b_{ik}$, and

$$c_{ik} = \frac{\Gamma(a_{ik} + b_{ik})}{\Gamma(a_{ik})\,\Gamma(b_{ik})} (x_{0k} - x_{-1k})$$

for $i = -1, 0$. It may be advantageous to keep $x_{0k} - x_{-1k}$ constant or nearly constant across investigators because the distance contributes to the weighted posterior. In addition, the values of (a_{ik}, b_{ik}) may be highly influential when they are not small. For example, if $(a_{i1}, b_{i1}) = (1, 2)$ and $(a_{i2}, b_{i2}) = (2, 4)$ with the same value of $x_{0k} - x_{-1k}$, then $c_{i2} \,/\, c_{i1} = 10$.

Here we provide the precise definition of the MTD. For a given risk level $\gamma \in (0, 1)$, we define $D_\gamma(\vec{\beta}) \equiv D_\gamma$ to be the dose that achieves $\pi_{D_\gamma} = \gamma$. Under the logistic model, we have

$$D_\gamma(\vec{\beta}) = \frac{\log\left(\frac{\gamma}{1-\gamma}\right) - \beta_0}{\beta_1}. \tag{2.12}$$

This is sometimes called the lethal dose γ. Assuming both efficacy and toxicity of a treatment are monotonic with respect to dose, D_γ, is the target treatment dose for a fixed risk level, and it is the parameter of interest. A typical value of γ is between 0.15 and 0.35 in dosing trials for serious medical conditions such as cancer.

2.4.2 Allocating Trial Participants to New Experimental Doses

We modify the patient gain discussed in Whitehead and Williamson [68] to the *patient loss* by taking the inverse of the gain. For the patient loss, we allocate the $(n+1)^{\text{th}}$ patient at

$$x_{n+1} = \text{argmin}_x \, E\left((x - D_\gamma)^2 \mid \vec{y}_n\right).$$

Thus, for the posterior weighted estimator, the next patient would be allocated at

$$\hat{D}^*_\gamma = \sum_{k=1}^{K} w_k \, \hat{D}_{\gamma,k}$$

where $\hat{D}_{\gamma,k} = E\left(D_\gamma \mid \vec{y}_n, Q_k\right)$ and $w_k = P(Q_k)$. Using the consensus prior estimator, the next patient would be allocated at

$$\hat{D}^{**}_\gamma = \sum_{k=1}^{K} w^*_k \, \hat{D}_{\gamma,k}\,.$$

where the k^{th} posterior mean contributes to this decision with the data-dependent weight of $w^*_k = P(Q_k \mid \vec{y}_n)$.

It is worth noting that due to the nonlinear transformation of the regression parameters involved in computation of the MTD, it is difficult to construct a self-consistent estimator

from the class

$$\hat{D}^*_\gamma = \sum_{k=1}^{K} w_k \, \hat{D}_{\gamma,k}$$

with the data-independent weight $w_k = P(Q_k)$.

2.4.3 Simulation Study

Simulation Design

We designed simulation studies to investigate the relative operating characteristics of the weighted posterior estimator and the consensus prior estimator in Phase I clinical trials. For numerical experiments, we set the target risk level at $\gamma = 0.2$ and assume that a maximum of $N = 25$ patients are to be treated, a common value for a phase I clinical trials. We suppose three investigators have divergent priors. The first investigator specifies Q_1 with $(x_{-11}, a_{-11}, b_{-11}) = (-4.0, 1.2, 3.8)$ and $(x_{01}, a_{01}, b_{01}) = (4.0, 1.2, 3.8)$. This prior specification added 1.2 AEs and 3.8 non-AEs at the arbitrarily low dose $x_{-11} = -4$ and added 3.8 AEs and 1.2 non-AEs at the arbitrarily high dose $x_{01} = 4$. The second investigator specifies Q_2 by $(x_{-12}, a_{-12}, b_{-12}) = (0.0, 1.2, 3.8)$ and $(x_{02}, a_{02}, b_{02}) = (8.0, 1.2, 3.8)$, and the third investigator specifies Q_3 by $(x_{-12}, a_{-12}, b_{-12}) = (4.0, 1.2, 3.8)$ and $(x_{02}, a_{02}, b_{02}) = (12.0, 1.2, 3.8)$. The three prior elicitations are equally strong by matching $x_{0k} - x_{-1k} = 8$ and $\sum_{i=-1}^{0} a_{ik} + b_{ik} = 10$. In addition, we assume equal prior probabilities so that $P(Q_k) = 1/3$ for $k = 1, 2, 3$. Note that Q_1 is relatively conservative and Q_3 is relatively anti-conservative. In particular, the prior means of D_γ are $D^*_{\gamma,k} = E(D_\gamma \mid Q_k) = -3.1, 0.9, 4.9$ for $k = 1, 2, 3$, respectively. If the trial were to proceed using each investigator separately, the first trial patient would be allocated to these three doses accordingly and remaining trial patients would be allocated using the updated estimates of D_γ. Therefore the three priors would

lead to different estimated doses and different numbers of AEs. On the other hand, if the three investigator opinions are incorporated into a single trial, the first patient is allocated at $(-3.1 + 0.9 + 4.9) / 3 = 0.9$ and the remaining decisions are based on the choice of estimator. If the weighted posterior estimator, $\hat{D}^*_\gamma$, is used, the three investigators contribute to the entire trial equally. If the consensus prior estimator, $\hat{D}^{**}_\gamma$, is used the contribution of each investigator adaptively varies depending on accumulated data.

As shown in Figure 2.2, we consider the nine scenarios by 3×3 possible pairing values of $\beta_0 = (-6, -3, 0)$ and $\beta_1 = (0.5, 0.8, 1.2)$ under the logistic regression model. The scenario numbers are assigned relative to the ascending order of D_γ. From Scenarios 1 to 9, the nine values of D_γ are -2.77, -1.73, -1.16, 1.34, 2.02, 3.23, 3.84, 5.77, and 9.23. A low value of D_γ implies that a new experimental dose is toxic relative to the dosing range, and a high value of D_γ implies that it is relatively toxic safe. In each subfigure, the true D_γ of each scenario is indicated by the dotted line, and prior guesses $D^*_{\gamma,k}$ for $k = 1, 2, 3$ are indicated by the three solid lines.

Evaluations

We let $\hat{D}_\gamma$ generically denote a Bayes estimator of D_γ. Then, the probability distribution of interest is the probability of an AE at dose $\hat{D}_\gamma$, $\pi_{\hat{D}_\gamma}$, rather than the distribution of $\hat{D}_\gamma$ itself. In other words, if $\hat{D}_\gamma$ deviates from the true D_γ, practitioners' interest is the degree of deviation from the true risk level.

We evaluate simulation results in three aspects. First, it is important to treat trial participants near the target risk level $\gamma = 0.2$ (from the perspective of current patients). In this regard, we measure the mean, standard deviation (SD), and RMSE of π_X with respect to 0.2, where X is the random variable denoting each treated dose during the trial. Second, it is also important to estimate D_γ when γ is near 0.2 at the end of a trial (from the perspective of

future patients). In this regard, we measure the mean, SD, and RMSE of $\pi_{\hat{D}}$ after observing the outcomes of $N = 25$ patients. Third, practitioners are generally concerned about the distribution $S_N = \sum_{i=1}^{N} Y_i$, the sum of AEs at the end of a trial. In particular, as we fixed $N = 25$ and $\gamma = 0.2$, a desirable distribution of S_{25} should have a mode near $N \times \gamma = 5$ with small $Var(S_{25})$ and large $P(4 \leq S_{25} \leq 6)$. To this end, we summarize the distribution of S_{25} by the mean, SD, and $P(4 \leq S_{25} \leq 6)$.

Simulation Results

The simulation results for separate trials using Q_1, Q_2, Q_3, individually, and a single trial incorporating $Q = \{Q_1, Q_2, Q_3\}$ are summarized in Table 2.1. In the table, results utilizing the weighted posterior estimator (i.e. $\hat{D}^*_\gamma$) are denoted by Q^*_{123}, and results utilizing the consensus prior estimator (i.e. $\hat{D}^{**}_\gamma$) are denoted by Q^{**}_{123}.

Focusing on the distribution of π_X, of the probability of an AE at each allocated dose level during the trial, results can be very sensitive when a single prior is used (see the left three columns of Table 1). In Scenario 1, Scenario 2, and Scenario 3, when a new experimental dose has relatively high toxicity, the respective means of π_X were 0.711, 0.787, and 0.860 for the anti-conservative prior Q_3 which imply that large portions of trial patients were overdosed in simulated trials. By the use of a consensus prior, the respective means reduced to 0.428, 0.393, and 0.363 for Q^*_{123}, and they reduced even greater to 0.268, 0.238, and 0.224 for Q^{**}_{123}, respectively, which were closer to $\gamma = 0.2$. The SDs of π_X were greater in Q^{**}_{123} than in Q^*_{123}. The RMSEs were lower for Q^{**}_{123}. When the true D_γ increases from Scenario 4 to Scenario 9, Q^{**}_{123} yielded the averages of π_X closer to $\gamma = 0.2$ with larger SDs when compared to Q^*_{123}. As a consequence, the RMSE of π_X was greater in Q^{**}_{123} than in Q^*_{123} from Scenario 4 to Scenario 7.

Focusing on the distribution of $\pi_{\hat{D}_\gamma}$ (see the middle three columns of the table), similar trends

were found for Q_1, Q_2, Q_3, Q^*_{123}, and Q^{**}_{123} with respect to the mean, SD, and RMSE. In each scenario, $\pi_{\hat{D}_\gamma}$ in Q^{**}_{123} was closer to the target $\gamma = 0.2$ on average with larger variability when compared to Q^*_{123}. The resulting RMSEs of Q^{**}_{123} were smaller in Scenarios 1, 2, 3, 8, and 9, the cases when the combined priors were relatively distant from the true D_γ. The gross negative impact of prior misspecification was reduced by weighting each prior based on data in the consensus prior estimator. When the combined priors were relatively close to the truth as in Scenarios 4, 5, 6, and 7, the RMSEs of Q^*_{123} were smaller due to less variability of $\hat{D}^*_\gamma$ than the variability of $\hat{D}^{**}_\gamma$. This general tendency was similar to the case under the normal model when we compared the self-consistent weighted posterior estimator $\hat{\mu}^*$ and the consensus prior estimator $\hat{\mu}^{**}$.

We turn our focus to the distribution of S_{25} (see the right three columns of the table). The resulting $E(S_{25})$ generally decrease as the true D_γ increases for each method of Bayes estimation with rare exceptions. In the nine scenarios, the range of $E(S_{25})$ was (0.066, 4.835) for Q_1, (0.433, 10.674) for Q_2, (2.044, 17.778) for Q_3, (0.436, 10.674) for Q^*_{123}, and (1.590, 6.686) for Q^{**}_{123}. By adaptively weighting each individual's opinion based on accumulated information, Q^{**}_{123} resulted in the shortest range of $E(S_{25})$ across the various scenarios, and it yielded S_{25} closest to $N \times \gamma = 5$ on average among the five considered approaches. Furthermore, the adaptive weighting scheme of the consensus prior provided robustness to $P(4 \leq S_{25} \leq 6)$.

In summary, the resulting distributions of π_X, $\pi_{\hat{D}_\gamma}$, and S_N from the individual priors Q_1, Q_2, and Q_3 showed high prior sensitivity, and robustness was gained by the use of a compromised design incorporating all priors. When we compared the weighted posterior estimator Q^*_{123} and the consensus prior estimator Q^{**}_{123}, Q^*_{123} yielded relatively small variability with respect to dose allocation and final estimation of D_γ, while Q^{**}_{123} yielded relatively small deviation from the target MTD on average. Further, by adjusting the weights of contributions to sequential decisions based on updated data, Q^{**}_{123} provided greater robustness with respect to the distribution of S_{25}.

Figure 2.3 graphically summarize the relative performance of Q_{123}^{**} to Q_{123}^{*}. Figure 2.3a plots the relative RMSE of π_X (solid) and the relative RMSE of $\pi_{\hat{D}_\gamma}$ (dotted). A relative RMSE lower than one implies a smaller RMSE for Q_{123}^{**}, and this subfigure is analogous to Figure 2.1a. When the consensus prior guess and the true value of a parameter of interest are fairly close, the design utilizing the weighted posterior estimator tends to perform better with respect to RMSE. However, when the two quantities are distant (i.e. one extreme prior guess is relatively close to the true value), the design utilizing the consensus prior estimator tends to perform better with respect to RMSE. Figure 2.3b highlights the robustness and practical importance (i.e. safety of dose-finding studies) of the design utilizing the consensus prior estimator with respect to $E(S_{25})$.

2.5 Discussion

In small-sample studies it is often necessary to incorporate prior information in order to stabilize estimation and inferential procedures. When several subject-specific experts are available for prior elicitation, the incorporation of multiple priors can lead to robustness to prior mis-specification when compared to the use of a single expert opinion. It is highly unlikely that two or more researchers have the exactly same prior information about a parameter of interest. For some scientific topics, experts have various and perhaps strong opinions. To this end, the incorporation of multiple priors via a consensus estimator can be useful in a practical sense. We have considered the performance of two classes of consensus estimators. First is the class of convex combinations of Bayes estimators that are derived from each individual prior, with weights for each Bayes estimator resulting from the weight placed on the corresponding prior distribution giving yield to that estimator. In the second class of estimators, a single consensus prior is first formed as a combination of the available priors, and the resulting Bayes estimator is then obtained. We have provided an alternative

theoretical perspective of the consensus prior estimator by showing that this class of estimators results in a natural posterior weighting scheme with prior weights updated conditional upon observed data.

The first class of weighted posterior estimators have a path to self-consistency under simple models, while the second class of consensus prior estimators does not generally result in self-consistency. However, the adaptive weighting scheme of the consensus prior estimator allows for adjustment of the contribution of each investigator's opinion based on empirical evidence, a desirable feature in an adaptive experiment. Based on our numerical illustrations, the weighted posterior estimator tends to have smaller variability while the consensus prior estimator tends to have smaller bias. If the true value of the parameter of interest is located near the mean of the prior guesses, use of weighted posterior estimator seems preferable. If this is not the case, use of the consensus prior estimator is generally preferable. The former is not Bayesian, the latter is.

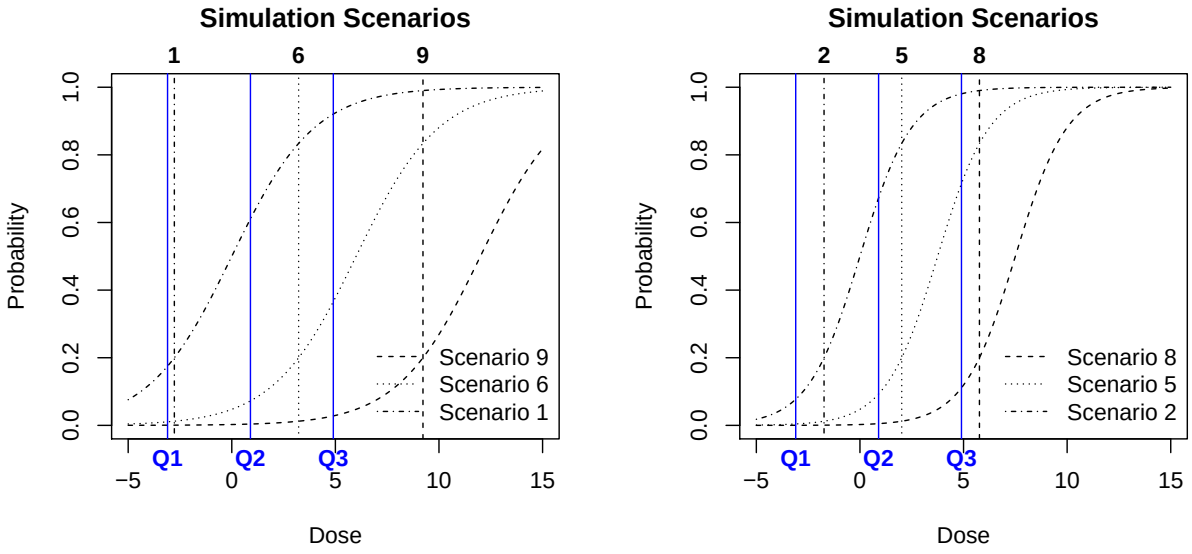
A primary focus of this chapter has been on the utility of incorporating multiple expert opinions into scientific analyses, when they are available, not to persuade the reader that the consensus estimator for compromising priors should always be chosen. Instead, choosing the consensus estimator should depend on the context of the scientific problem. For example, in Scenario 7 of Table 2.1, the scheme utilizing the consensus prior estimator allocated patients at the risk level of 0.181 on average with SD of 0.144. On the other hand, the scheme utilizing the weighted posterior estimator allocated patients at the risk level of 0.103 on average with SD of 0.086 which resulted in the smaller RMSE (0.145 versus 0130). To make an informed decision, one must consider whether precise underdosing throughout a trial or slightly less precise dosing near the target is more preferable.

As a general precaution, when one formulates a set of priors, it is important to check the balance of prior elicitations. Specifically, one will probably not want an investigator's prior to be upweighted simply because it is dogmatic. To this end, it may be necessary to impose

some restriction on the magnitude of hyper-parameters. In general, matching the hyper-parameters across investigators and investigating the impact of different hyper-parameters under the model specified is strongly encouraged unless intentional differential weight to investigators is desired.

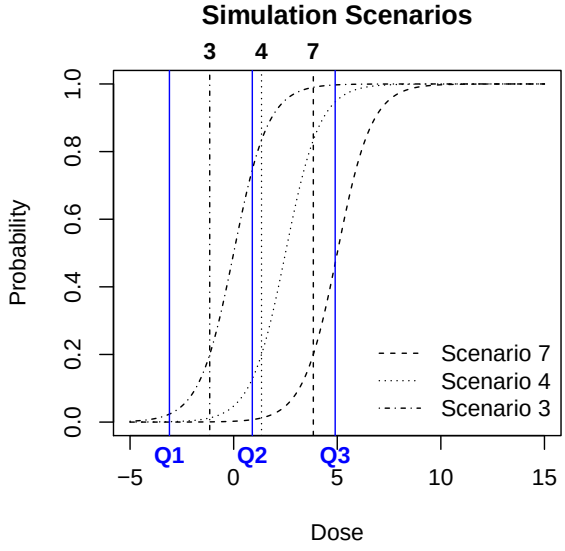
While further work on evaluating the performance of consensus estimators should be performed, the incorporation of multiple priors into analyses can provide robustness to prior misspecification when compared to the use of a single expert opinion.

Figure 2.2: Simulation scenarios for Phase I clinical trials



(a) Scenarios 9, 6 and 1

(b) Scenarios 8, 5 and 2

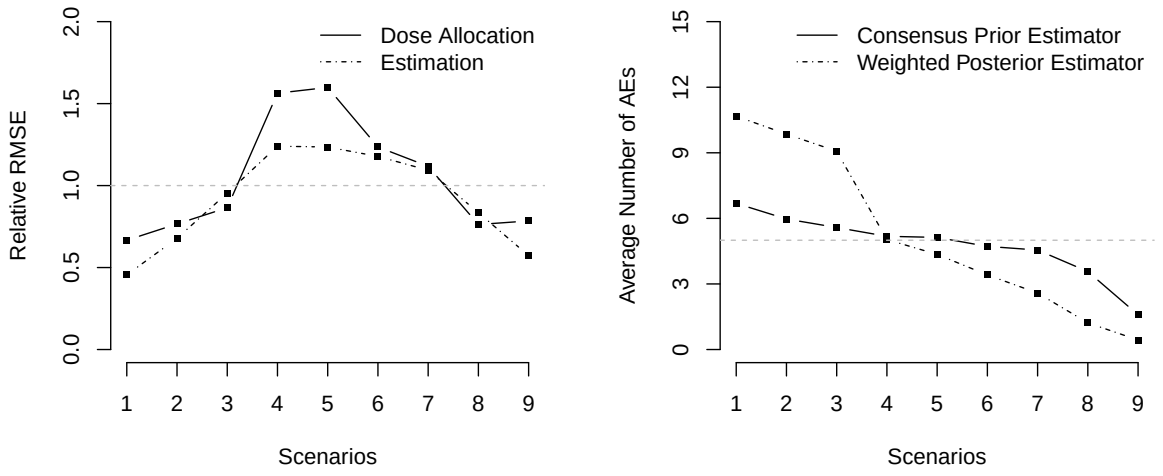


(c) Scenarios 7, 4 and 3

Table 2.1: Simulation results

Scenario	Prior	Dose Allocations			MTD Estimation			Number of AEs		
		Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	$P(4 \leq S_{25} \leq 6)$
Scenario 1	Q_1	0.194	0.071	0.071	0.195	0.067	0.067	4.835	1.108	0.831
	Q_2	0.363	0.107	0.195	0.282	0.046	0.094	9.077	1.549	0.033
	Q_3	0.711	0.075	0.516	0.646	0.006	0.446	17.778	2.039	0.000
	Q^*_{123}	0.428	0.089	0.245	0.365	0.041	0.170	10.674	1.706	0.003
	Q^{**}_{123}	0.268	0.149	0.163	0.211	0.077	0.077	6.686	1.212	0.450
Scenario 2	Q_1	0.159	0.086	0.095	0.186	0.080	0.081	3.992	0.961	0.688
	Q_2	0.316	0.140	0.182	0.228	0.065	0.071	7.891	1.228	0.117
	Q_3	0.787	0.077	0.592	0.717	0.006	0.517	19.676	1.882	0.000
	Q^*_{123}	0.393	0.122	0.228	0.318	0.057	0.131	9.856	1.460	0.005
	Q^{**}_{123}	0.238	0.171	0.175	0.196	0.089	0.089	5.971	1.018	0.714
Scenario 3	Q_1	0.141	0.103	0.119	0.187	0.093	0.094	3.554	0.861	0.510
	Q_2	0.279	0.166	0.184	0.197	0.080	0.080	7.021	0.976	0.300
	Q_3	0.860	0.064	0.663	0.795	0.004	0.595	21.490	1.637	0.000
	Q^*_{123}	0.363	0.153	0.223	0.283	0.073	0.110	9.072	1.205	0.008
	Q^{**}_{123}	0.224	0.192	0.193	0.196	0.105	0.105	5.582	0.877	0.861
Scenario 4	Q_1	0.052	0.056	0.159	0.139	0.074	0.096	1.273	0.725	0.002
	Q_2	0.186	0.111	0.112	0.192	0.094	0.094	4.652	0.879	0.907
	Q_3	0.418	0.190	0.289	0.275	0.052	0.091	10.442	1.409	0.000
	Q^*_{123}	0.200	0.106	0.106	0.201	0.088	0.088	5.032	0.894	0.928
	Q^{**}_{123}	0.208	0.165	0.166	0.202	0.109	0.109	5.179	0.896	0.916
Scenario 5	Q_1	0.044	0.037	0.160	0.103	0.043	0.106	1.111	0.779	0.002
	Q_2	0.167	0.087	0.093	0.187	0.080	0.081	4.185	0.966	0.756
	Q_3	0.333	0.144	0.196	0.238	0.060	0.072	8.298	1.278	0.066
	Q^*_{123}	0.174	0.081	0.085	0.191	0.074	0.075	4.350	0.975	0.801
	Q^{**}_{123}	0.205	0.136	0.136	0.201	0.092	0.092	5.129	1.050	0.861
Scenario 6	$Q1$	0.041	0.024	0.161	0.077	0.024	0.125	1.020	0.819	0.003
	Q_2	0.134	0.059	0.089	0.167	0.062	0.070	3.362	1.051	0.429
	Q_3	0.262	0.087	0.107	0.221	0.064	0.067	6.551	1.248	0.494
	Q^*_{123}	0.137	0.055	0.084	0.165	0.057	0.067	3.413	1.065	0.442
	Q^{**}_{123}	0.188	0.103	0.103	0.200	0.079	0.079	4.717	1.154	0.797
Scenario 7	$Q1$	0.005	0.005	0.196	0.020	0.004	0.180	0.113	0.325	0.000
	Q_2	0.107	0.092	0.131	0.180	0.091	0.093	2.676	0.808	0.140
	Q_3	0.234	0.137	0.141	0.192	0.089	0.090	5.840	0.893	0.785
	Q^*_{123}	0.103	0.086	0.130	0.168	0.089	0.095	2.584	0.770	0.106
	Q^{**}_{123}	0.181	0.144	0.145	0.197	0.104	0.104	4.552	0.889	0.883
Scenario 8	$Q1$	0.003	0.003	0.197	0.010	0.001	0.190	0.083	0.280	0.000
	Q_2	0.051	0.041	0.155	0.113	0.049	0.100	1.263	0.802	0.005
	Q_3	0.174	0.088	0.092	0.190	0.082	0.082	4.372	0.985	0.802
	Q^*_{123}	0.049	0.040	0.156	0.108	0.049	0.104	1.221	0.777	0.003
	Q^{**}_{123}	0.142	0.104	0.119	0.190	0.086	0.087	3.567	0.950	0.508
Scenario 9	$Q1$	0.003	0.002	0.197	0.006	0.001	0.194	0.066	0.253	0.000
	Q_2	0.018	0.010	0.183	0.037	0.008	0.163	0.436	0.608	0.000
	Q_3	0.081	0.042	0.126	0.126	0.046	0.087	2.044	0.966	0.068
	Q^*_{123}	0.018	0.010	0.183	0.036	0.008	0.164	0.433	0.595	0.000
	Q^{**}_{123}	0.064	0.044	0.143	0.118	0.045	0.094	1.590	0.887	0.018

Figure 2.3: Relative RMSE and average number of AEs



(a) Relative RMSE

(b) Average Number of AEs

Chapter 3

Balancing Individual- and Population-Level Ethics in Phase I Trials

In Chapter 3, we turn our focus to ethical issues in Phase I clinical trials. Bayesian adaptive designs have been proposed for Phase I clinical trials since the continual reassessment method (CRM) was proposed by O'Quigley, Pepe and Fisher [45]. Focused on dose-finding in the setting of cancer chemotherapies, the CRM seeks to allocate new patients to the posterior estimate of a maximum tolerable dose (MTD). Later, Whitehead and Brunier applied Bayesian decision theory to maximize statistical information for the MTD when allocating new patients to an experimental dose [67]. The two allocation rules reflect conflicting perspectives of dose-finding trials. The CRM emphasizes individual-level ethics concerned with treating current trial participants, while the method of Whitehead and Brunier emphasizes population-level ethics concerned with treating future patients at the most precise estimate possible. As a natural solution, we consider compromising between the two conflicting perspectives using a novel Bayesian decision theoretic dose allocation method.

3.1 Introduction

Early in the drug development process, data on the toxicity of a new experimental treatment is often scarce. The primary objective of a Phase I clinical trial is to study the toxicity of an experimental treatment and to determine a maximum tolerable dose (MTD) for future patients. Because little information regarding toxicity is available, over- and underdosing patients is inevitable in these early trials. In addition, due to small sample sizes generally considered in the first phase of drug development, resulting estimates of the MTD often suffer from low precision.

To address many of the ethical and statistical concerns often raised during the conduct of dose-finding trials, O'Quigley, Pepe and Fisher [45] proposed the continual reassessment method (CRM) for Phase I cancer studies. The CRM is a Bayesian adaptive design that utilizes a prior distribution for the dose-response curve, together with all available observed data, when allocating the next randomized patient to a new treatment dose. This design proceeds with a fixed allowable toxicity and a more precise definition of the MTD. If Y is a binary random variable with $Y = 1$ indicating an AE, we define MTD_γ as the dose which satisfies $P(Y = 1 \mid \text{MTD}_\gamma) = \gamma$ for a fixed toxicity level γ. The CRM then seeks to treat each new trial patient at MTD_γ, hence dose escalation and de-escalation naturally occur by the amount of change in posterior estimation of MTD_γ.

Since the CRM was introduced many modifications of the design have been proposed to reduce overdosing patients and to meet various practical needs. Goodman, Zahurak and Piantadosi [29] suggested assigning more than one patient at a time to a given dose and to limit dose escalation by a single prefixed dose level. Their simulation studies showed that this restricted rule effectively reduced the number of observed AEs on average. Babb, Rogatko and Zacks [3] proposed an escalation with overdose control (EWOC) design that suggested treating a new patient at a fixed quantile of the marginal posterior distribution

of MTD_γ. Relative to the CRM, the EWOC imposes a heavier penalty for overdosing than underdosing. Other authors have considered sensitivity to the choice of prior distribution in these algorithms as well [3].

While the CRM and various modifications have primarily emphasized individual-level ethics by allocating new patients at or below the posterior mean of the MTD, Whitehead and Brunier [67] considered an approach to provide greater emphasis on population-level ethics. Specifically, Whitehead and Brunier [67] applied Bayesian decision theory to Phase I clinical trials by devising a dose allocation scheme from the perspective of future treated patients. In this approach a gain function was defined as the inverse of the asymptotic variance of the maximum likelihood estimate (MLE) for MTD_γ. The gain function seeks to optimize dose allocation in order to maximize precision of the resulting estimate of MTD_γ. As such, this design is referred to as the information gain method (IGM) in this chapter.

The IGM was motivated by a trial setting with healthy volunteers and where potential AEs from a new drug are characterized by mild to moderate reactions (e.g. headache, drowsiness, rash, etc.). In contrast, the applications given in O'Quigley, Pepe and Fisher [45] included radiolabeled tumor-specific monoclonal antibodies and combination chemotherapies which are relatively more serious cases. As depicted in their applications, the ethical perspective in Phase I clinical trials should depend upon the severity of treating diseases and potential AEs. Whitehead and Williamson [68] later discussed additional gain functions in cancer therapy settings and pointed out that conservatively modified CRMs may reduce the speed of trial procedures but also noted that the constraints suggested by Faries [27] and Goodman, Zahurak and Piantadosi [29], starting at the lowest available dose level and prohibiting skipping a fixed dose level, may not benefit patients in the early stage of a trial. Furthermore, such constrains slow down the rate of information growth.

Additional Bayesian designs focusing on efficient estimation of the MTD have been discussed by Haines, Perevozskaya and Rosenberger [31] including so called c-optimal designs and D-

optimal designs. However, O'Quigley and Conaway [44] pointed out that under current guidelines it is difficult to justify utilizing a procedure that fully sacrifices the point of view of current trial participants for that of future patients. Clearly, both precise estimation of the MTD and ethical dose allocation are of paramount concerns in early phase trials. As such, a natural solution to the two conflicting perspectives would be a balance between individual- and population-level ethics. Bartroff and Lai [5] proposed a method that minimizes a loss function consisting of two additive terms, where each term reflects the individual and population risks. This approach attempted to compromise between the loss of a new patient and the loss of the following patient who will receive a dose that minimizes the following patient's loss. Due to the complexity that arises from considering infinitely many future patients they chose to focus on the loss associated with two additional future patients.

Recently Oron and Hoff [47] argued that up-and-down designs, such as the traditional 3+3 design [60], offer more robust results with respect to the number of cohorts treated at the MTD when compared to the CRM. They pointed out that the CRM is sensitive to prior specification and settles on a specific dose which is not necessarily the MTD in small samples. Though the simple up-and-down designs were not originally developed based on a rigorous statistical framework, they attempt to gather information regarding toxicity over a relatively wide dose range. In this sense, the comparisons between the CRM and up-and-down schemes can be viewed as a simple version of individual- versus population-level ethics as well.

In the design of a Phase I dose-finding trial to investigate hyperthermic intraperitoneal chemotherapy (HIPEC) we sought to compromise two well-known and commonly used dose allocation methods (CRM and IGM) to emphasize individual-level ethics while increasing the precision of the resulting MTD estimate. To this end, we propose an adaptive design to balance the two perspectives using a new method we refer to as the balanced information gain method (BIGM). The BIGM is formally developed by decomposing the loss function for the IGM and observing that the IGM naturally compromises individual- and population-

level ethics, but that the relative weight allocated to each attribute results in high volatility in dose allocations during the early phase of a trial. To account for this, we modify the loss function with a tuning parameter that allows a trialist to differentially weigh individual- and population-level ethics to meet the necessary goals of their particular clinical setting. In Section 3.2 we detail the statistical methodology used in developing the BIGM. Section 3.3 presents simulation studies to illustrate the operating characteristics of the newly proposed approach, and Section 3.4 presents selected results from our experience in using the proposed method to evaluate designs for a dosing trial of hyperthermic intraperitoneal chemotherapy (HIPEC). Section 3.5 concludes with a general discussion of the BIGM and its use in early phase dosing studies.

3.2 Methods

A Bayesian decision theoretic approach to adaptive design can be thought of as a sequence of decisions based upon an action space, a loss (or gain) function, a likelihood function, and a prior distribution. In the context of a fully sequential phase I dosing algorithm, each new subject is treated at the dose that minimizes expected loss (maximizes expected gain), which in turn is a function of the assumed probability model and prior distribution. In this section we develop the BIGM as a compromise between the CRM and IGM.

3.2.1 Decomposition of the Loss Function in the IGM

The IGM is a dose allocation scheme in which the loss function is taken to be the asymptotic variance of the MLE for MTD_γ as estimated via a generalized linear model (GLM). In the context of a GLM, let $\eta = \beta_1 + \beta_2 x$, $\beta_2 > 0$, denote a linear predictor and $F(\eta)$ denote a distribution function representing the probability of an AE at dose x (in log-scale).

Assuming independent binary random variables with probability $F(\eta_i)$, where $\eta_i = \beta_1 + \beta_2 x_i$ for $i = 1, \ldots, n$, the likelihood function for $\vec{\beta} = (\beta_1, \beta_2)$ is

$$L_n \equiv f(\vec{y}_n \mid \vec{\beta}) = \prod_{i=1}^{n} \left[F(\eta_i)\right]^{y_i} \left[1 - F(\eta_i)\right]^{1-y_i} \ .$$

Denoting the log-likelihood function by $l_n \equiv \log(L_n)$, Fisher's expected information matrix based upon $\vec{y}_n$ and $\vec{x}_n = (x_1, \ldots, x_n)$ is given by

$$\mathcal{I}_n \equiv -E\left(\frac{\partial^2 l_n}{\partial\vec{\beta}\,\partial\vec{\beta}^T}\right) = \begin{pmatrix} \sum_{i=1}^n \tau_i & \sum_{i=1}^n \tau_i x_i \\ \sum_{i=1}^n \tau_i x_i & \sum_{i=1}^n \tau_i x_i^2 \end{pmatrix} ,$$

where

$$\tau_i \equiv \frac{\left[\partial F(\eta_i) \,/\, \partial\eta_i\right]^2}{F(\eta_i)\left[1 - F(\eta_i)\right]}$$

can be thought of as an unnormalized weight corresponding to the i^{th} observation at dose x_i.

Denoting the inverse of $F(\cdot)$ by $F^{-1}(\cdot)$, MTD_γ in the above model specification is given by the transformation h: $\mathcal{R}^2 \to \mathcal{R}$ with

$$h(\vec{\beta}) = \frac{F^{-1}(\gamma) - \beta_1}{\beta_2} \equiv \text{MTD}_\gamma \, ,$$

where the gradient of h is given by

$$\nabla\vec{h} \equiv \frac{\partial h(\vec{\beta})}{\partial\vec{\beta}} = -\begin{pmatrix} \frac{1}{\beta_2} & \frac{F^{-1}(\gamma)-\beta_1}{\beta_2^2} \end{pmatrix}^T = -\frac{1}{\beta_2}\begin{pmatrix} 1 & \text{MTD}_\gamma \end{pmatrix}^T \ .$$

Further denoting the normalized weight from the i^{th} observation by $w_i \equiv \tau_i \,/\, \sum_{i=1}^n \tau_i$, the

asymptotic variance of $\widehat{\text{MTD}}_\gamma$, the MLE of MTD_γ, is

$$\begin{aligned}
\nabla\vec{h}^T\mathcal{I}_n^{-1}\nabla\vec{h} &= \frac{(\text{MTD}_\gamma)^2\left(\sum_{i=1}^n \tau_i\right) - 2\,\text{MTD}_\gamma\left(\sum_{i=1}^n \tau_i x_i\right) + \sum_{i=1}^n \tau_i x_i^2}{\beta_2^2\det(\mathcal{I}_n)} \\
&= \frac{\left(\sum_{i=1}^n \tau_i\right)\left\{(\text{MTD}_\gamma)^2 - 2\,\text{MTD}_\gamma\left(\sum_{i=1}^n w_i x_i\right) + \sum_{i=1}^n w_i x_i^2\right\}}{\beta_2^2\left\{\left(\sum_{i=1}^n \tau_i\right)\left(\sum_{i=1}^n \tau_i x_i^2\right) - \left(\sum_{i=1}^n \tau_i x_i\right)^2\right\}} \\
&= \frac{\left(\text{MTD}_\gamma - \sum_{i=1}^n w_i x_i\right)^2 + \left\{\sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i\right)^2\right\}}{\beta_2^2\left(\sum_{i=1}^n \tau_i\right)\left\{\sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i\right)^2\right\}}.
\end{aligned} \tag{3.1}$$

In Equation (3.1), we let $m_n^{(1)} \equiv \sum_{i=1}^n w_i x_i \equiv \mu_n$ and $m_n^{(2)} \equiv \sum_{i=1}^n w_i x_i^2$ denote the first two weighted sample moments for the observed doses $\vec{x}_n = (x_1, \ldots, x_n)$. Then, the numerator represents the sum of squared error, $(\text{MTD}_\gamma - \mu_n)^2$, and sampling variance, $\sigma_n^2 \equiv m_n^{(2)} - \mu_n^2$, while the denominator consists of σ_n^2 which is equivalent to the gain function in D-optimal designs. From Equation (3.1), one can distinguish two main factors contributing to the variance of $\widehat{\text{MTD}}_\gamma$: the distribution of $\vec{x}_n$ and distances of $\vec{x}_n$ from the target estimand MTD_γ. The later source of the variance is straightforward as it appears in the numerator only, but the former source is rather complicated as it appears in both the numerator and denominator with a nondecreasing quantity $\sum_{i=1}^n \tau_i$ in the denominator. Equation (3.1) suggests that it is essential to increase σ_n^2 (i.e. a wide sampling scheme) to effectively reduce the variance of $\widehat{\text{MTD}}_\gamma$ when n is small (i.e. early phase of a trial), but such a wide sampling scheme is not necessarily optimal for the variance reduction when n is large. Instead, dose allocations near the true MTD_γ become more effective for variance reduction when n is large.

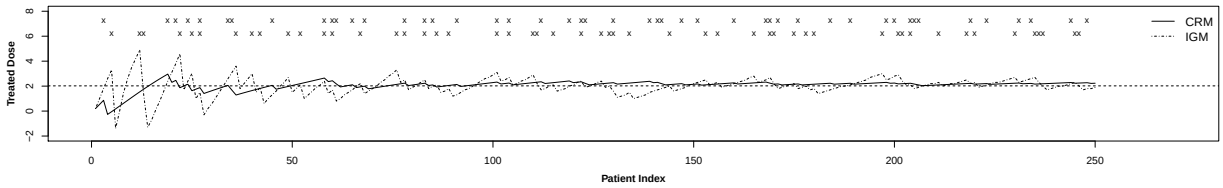
When deciding on an optimal dose for a new patient, it is necessary to consider $\mathcal{I}_{n+1}$ instead of $\mathcal{I}_n$ and further decompose the loss function $\mathcal{L}_I(x_{n+1} \mid \vec{\beta}) \equiv \nabla\vec{h}^T\mathcal{I}_{n+1}^{-1}\nabla\vec{h}$. For concise notation, let $z_n^{(m)} = \sum_{i=1}^n \tau_i x_i^m$ for $m = 0, 1, 2$. Then, it can be shown that the loss function with respect to a new dose x_{n+1} for given parameter $\vec{\beta}$ is decomposed as

$$\mathcal{L}_I(x_{n+1} \mid \vec{\beta}) \propto \frac{\tau_{n+1}\left(x_{n+1} - \text{MTD}_\gamma\right)^2 + z_n^{(0)}\left\{(\text{MTD}_\gamma - \mu_n)^2 + \sigma_n^2\right\}}{\left\{z_n^{(0)} z_n^{(2)} - (z_n^{(1)})^2\right\} + z_n^{(0)}\tau_{n+1}\left\{(x_{n+1} - \mu_n)^2 + \sigma_n^2\right\}}. \tag{3.2}$$

A derivation of Equation (3.2) is provided in the Appendix A.2. The implication of Equation (3.2) is similar to that of Equation (3.1). The new dose x_{n+1} affects the first term $(x_{n+1} - \text{MTD}_\gamma)^2$ in the numerator and the second term $(x_{n+1} - \mu_n)^2$ in the denominator. In order to reduce the loss associated with x_{n+1}, $\mathcal{L}_I(x_{n+1} \mid \vec{\beta})$, we need to account for a balance between the individual-level loss $(x_{n+1} - \text{MTD}_\gamma)^2$ and the population-level gain $(x_{n+1} - \mu_n)^2$.

Analogous to Equation (3.1), a small increase in the population gain $(x_{n+1} - \mu_n)^2$ can effectively reduce the loss $\mathcal{L}_I(x_{n+1} \mid \vec{\beta})$ as $n \to \infty$ because the increasing quantity $z_n^{(0)}$ is attached multiplicatively to $(x_{n+1} - \mu_n)^2$. As such, the IGM itself attempts to balance the perspective of current patients against the perspective of future patients, but the rate of compromise is rather slow as shown in Figure 3.1. The figure is from a simulated trial of sample size $N = 250$ to compare general sampling schemes under the CRM and IGM in large samples. The sample size is unrealistically large for a Phase I clinical trial, but it provides a useful contrast. In the figure, the solid and dotted trajectories are representative sampling paths under the CRM and IGM (respectively), and the dotted horizontal line indicates the true MTD_γ in this simulated trial. The first and second rows of "x" marks indicate observed AEs in the CRM and the IGM, respectively. The starting dose is low because we set the prior expectation of MTD_γ to be lower than the true MTD_γ, and the CRM required the treatment of 12 patients before reaching the MTD_γ from below. For a typical sample size $20 \leq N \leq 30$, about half of the trial patients would be undertreated. The slow movement is also worrisome if the trial is initiated at a high dose. On the other hand, the IGM algorithm has a wider sampling scheme which becomes stable as the trial proceeds, and the distribution of $\vec{x}_n$ is more robust to the starting dose level. However, such a large amplitude of dose allocations in the early stage of a trial has not been widely accepted, particularly in cancer therapy studies where observed AEs tend to be severe.

Figure 3.1: Comparing sampling paths of CRM and IGM in a large sample



3.2.2 Balanced Information Gain Method (BIGM)

After decomposing the loss function of the IGM, a compromise between the individual- and population-level perspectives can be achieved in infinitely many ways. We propose a simple modification of the numerator as

$$\mathcal{L}_B(x_{n+1} \mid \vec{\beta}, \lambda) \propto \frac{k_n(\lambda)\,\tau_{n+1}\,(x_{n+1} - \mathrm{MTD}_\gamma)^2 + z_n^{(0)}\left\{(\mathrm{MTD}_\gamma - \mu_n)^2 + \sigma_n^2\right\}}{\left\{z_n^{(0)} z_n^{(2)} - (z_n^{(1)})^2\right\} + z_n^{(0)}\tau_{n+1}\left\{(x_{n+1} - \mu_n)^2 + \sigma_n^2\right\}}. \tag{3.3}$$

for some nondecreasing $k_n(\lambda) > 1$, where $\lambda > 0$ is a tuning parameter which adjusts a level of compromise. A proposed accelerating function $k_n(\lambda)$ is given by

$$k_n(\lambda) = \left(1 + \frac{n}{N}\right)^{\lambda\left(1+\frac{s_n}{N\gamma}\right)},$$

where $s_n = \sum_{i=1}^{n} y_i$ is the number of AEs observed in n patients. Since the proposed $k_n(\lambda) > 1$ is a nondecreasing function with respect to n, the loss function $\mathcal{L}_B(x_{n+1} \mid \vec{\beta}, \lambda)$ gradually approaches a CRM-like sampling scheme as a trial proceeds. By letting $\lambda \to 0$ and $\lambda \to \infty$, the behavior of $\mathcal{L}_B$ is similar to that for the loss function of the IGM and the CRM, respectively. Specifically, relative to the loss function used in IGM, $\mathcal{L}_B$ with $\lambda > 0$ imposes a stronger penalty on the deviation from MTD_γ as n increases and the exponent of $k_n(\lambda)$ increases for each observed AE. We refer to the dose allocation method that seeks to minimize this modified loss function as the *balanced information gain method* (BIGM).

3.3 Simulations

Early phase clinical trialists are often interested in both the distribution of the estimate of MTD_γ and the total number of observed AEs at the conclusion of a trial. Through simulation we investigate the operating characteristics of the CRM, the IGM, and the BIGM as we vary the tuning parameter λ from zero to a large value. Operating characteristics are assessed under various degrees of prior mis-specification relative to the true dose response model.

3.3.1 Setting

We assume the probability of an AE at dose x (in log-sclale) is given by

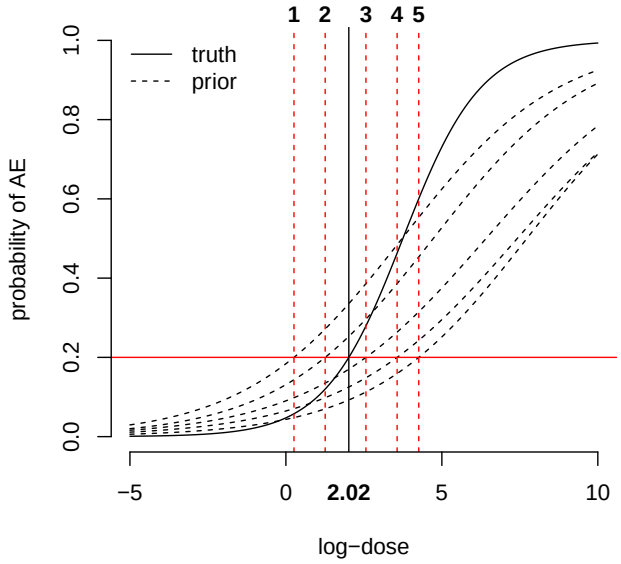
$$P(Y = 1 \mid x) = \frac{e^{-3.0+0.8x}}{1 + e^{-3.0+0.8x}}\,. \tag{3.4}$$

That is, the underlying dose-response curve is a logistic model with intercept parameter $\beta_1 = -3.0$ and slope parameter $\beta_2 = 0.8$. We fix the target risk level at $\gamma = 0.2$ yielding a true $\text{MTD}_{0.2}$ of 2.02. The action space is the set of real numbers in the CRM, and it is discretized by 0.1 for numerical search in the BIGM and the IGM. Each simulated trial is terminated when reaching a sample size of $N = 25$. Estimated operating characteristics are based upon 10,000 simulated trials for each considered scenario defined by the prior specification given in the next section.

3.3.2 Priors

For prior specifications we consider the method known as the conditional means prior [67, 6, 54], an interpretable and convenient approach for eliciting priors from scientific collaborators that serves as an attractive alternative to direct prior specification of the joint

Figure 3.2: Simulation setting



density function for (β_1, β_2). We let x_{-1} and x_0 be two arbitrary doses where the probabilities of an AE at the two doses are denoted by by θ_{-1} and θ_0, respectively. We specify the prior distribution for θ_i as $\theta_i \sim \text{Beta}(a_i, b_i)$ for $i = -1, 0$. Assuming independence between θ_{-1} and θ_0 and the corresponding Jacobian for the transformation, the joint prior density for $\vec{\beta} = (\beta_1, \beta_2)$ is induced as

$$f(\vec{\beta}) = \mathbf{1}_{\vec{\beta} \in \Omega} (x_0 - x_{-1}) \prod_{i=-1}^{0} [F(\eta_i)]^{a_i - 1} [1 - F(\eta_i)]^{b_i - 1} \frac{\partial F(\eta_i)}{\partial \eta_i} ,$$

where $\Omega \subset (-\infty, \infty) \times (0, \infty)$ is the support of $\vec{\beta}$. Using the logistic link, we obtain conjugacy as

$$f(\vec{\beta}) = \mathbf{1}_{\vec{\beta} \in \Omega} (x_0 - x_{-1}) \prod_{i=-1}^{0} [F(\eta_i)]^{a_i} [1 - F(\eta_i)]^{b_i} ,$$

and elicited a_i and b_i can be thought of as pseudo observations at the two doses.

To examine the performance of each procedure under varying degrees of prior mis-specification we consider five different priors. Priors 1 to 5 specify $(-2, 6)$, $(-1, 7)$, $(0, 9)$, $(1, 10)$ and $(2, 10)$ for (x_{-1}, x_0), respectively. We commonly let $(a_{-1}, b_{-1}) = (1.003, 1.303)$ for each low dose x_{-1} and $(a_0, b_0) = (1.303, 1.003)$ for each high dose x_0. The hyper-parameter $(a_{-1}, b_{-1}) = (1.003, 1.303)$ corresponds to the prior mode $\hat{\theta}_{-1} = 0.01$ and $P(\theta_{-1} < 0.9) = 0.95$ which reflects large uncertainty regarding the risk at each low dose x_{-1}. Similarly, $(a_0, b_0) = (1.303, 1.003)$ corresponds to $\hat{\theta}_0 = 0.99$ with $P(\theta_0 > 0.1) = 0.95$. In Figure 3.2 the five dotted dose-response curves are

$$P(Y = 1 \mid x) = \frac{e^{E(\beta_1)+E(\beta_2)\,x}}{1 + e^{E(\beta_1)+E(\beta_2)\,x}}$$

for each prior. The solid curve in the figure is the true dose-response curve in Equation (3.4). On top of the figure, each number indicates the point $\text{MTD}_{0.2} = \frac{\log(0.25)-E(\beta_1)}{E(\beta_2)}$ for each prior. We can see that prior estimate of $\text{MTD}_{0.2}$, and hence the starting dose at trial initiation, gradually increases from Prior 1 to Prior 5.

3.3.3 Results

Our proposed BIGM algorithm was developed as a continuous compromise between the IGM and the CRM, with the goal of inducing a CRM-like sampling scheme as a trial proceeds and accelerating after each observed AE. Example dosing paths of the IGM, the BIGM with $\lambda = 0.5, 1.0, 1.5, 2.0$, and the CRM are compared in Figure 3.3. For illustration, the sample paths were generated under the true model defined in Secton 3.1 and using Prior 1 as defined in Section 3.2 (resulting in a highly conservative starting dose). The figure illustrates that the IGM and BIGM, with low values of λ, allow for greater variability in the dose allocation scheme when treating early trial participants. This behavior can be beneficial relative to the CRM for misspecified priors that tend to start at severely low or high dosing levels, since

Table 3.1: Simulation results

IGM									
	Posterior Mean			**MLE**			**Number of AEs**		
Prior	**Bias (PB%)[A]**	**Variance**	**MSE**	**Bias (PB%)**	**Variance**	**MSE**	**Mean**	**Variance**	$P(4 \leq S_{25} \leq 6)$
1	-0.540 (-26.7)	0.432	0.723	0.077 (3.8)	0.563	0.569	4.755	0.298	1.000
2	-0.404 (-20.0)	0.388	0.551	0.082 (4.1)	0.568	0.568	5.061	0.255	0.997
3	-0.246 (-12.2)	0.344	0.405	0.090 (4.5)	0.570	0.570	5.458	0.300	0.980
4	-0.019 (-0.9)	0.266	0.266	0.110 (5.4)	0.576	0.576	5.922	0.381	0.876
5	0.278 (13.8)	0.148	0.225	0.066 (3.3)	0.700	0.700	6.610	0.694	0.513
BIGM $\lambda = 0.5$									
	Posterior Mean			**MLE**			**Number of AEs**		
Prior	**Bias (PB%)**	**Variance**	**MSE**	**Bias (PB%)**	**Variance**	**MSE**	**Mean**	**Variance**	$P(4 \leq S_N \leq 6)$
1	-0.472 (-23.4)	0.418	0.641	-0.002 (-0.1)	1.095	1.095	4.492	0.495	0.951
2	-0.347 (-17.2)	0.397	0.517	0.005 (0.2)	0.751	0.751	4.853	0.515	0.984
3	-0.167 (-8.3)	0.368	0.396	0.062 (3.1)	0.680	0.684	5.348	0.544	0.948
4	0.020 (1.0)	0.270	0.270	0.059 (2.9)	0.497	0.501	5.923	0.622	0.790
5	0.289 (14.3)	0.167	0.250	0.043 (2.1)	0.883	0.884	6.711	0.964	0.434
BIGM $\lambda = 1.0$									
	Posterior Mean			**MLE**			**Number of AEs**		
Prior	**Bias (PB%)**	**Variance**	**MSE**	**Bias (PB%)**	**Variance**	**MSE**	**Mean**	**Variance**	$P(4 \leq S_N \leq 6)$
1	-0.414 (-20.5)	0.451	0.623	0.007 (0.3)	1.762	1.762	4.366	0.669	0.866
2	-0.314 (-15.5)	0.429	0.527	-0.013 (-0.6)	4.294	4.294	4.786	0.699	0.943
3	-0.169 (-8.4)	0.381	0.409	0.017 (0.8)	0.573	0.573	5.375	0.704	0.913
4	0.022 (1.1)	0.286	0.287	0.034 (1.7)	0.631	0.632	5.987	0.828	0.730
5	0.278 (13.8)	0.173	0.250	0.024 (1.2)	1.696	1.696	6.870	1.183	0.383
BIGM $\lambda = 2.0$									
	Posterior Mean			**MLE**			**Number of AEs**		
Prior	**Bias (PB%)**	**Variance**	**MSE**	**Bias (PB%)**	**Variance**	**MSE**	**Mean**	**Variance**	$P(4 \leq S_N \leq 6)$
1	-0.392 (-19.4)	0.470	0.623	0.011 (0.5)	36.651	36.651	4.269	0.823	0.797
2	-0.280 (-13.9)	0.447	0.525	-0.038 (-1.9)	8.236	8.238	4.734	0.894	0.890
3	-0.160 (-7.9)	0.400	0.425	-0.017 (-0.8)	1.680	1.680	5.403	0.893	0.876
4	0.008 (0.4)	0.308	0.308	0.023 (1.1)	0.689	0.689	6.112	1.039	0.666
5	0.246 (12.2)	0.172	0.232	0.033 (1.6)	1.914	1.915	7.073	1.357	0.320
CRM									
	Posterior Mean			**MLE**			**Number of AEs**		
Prior	**Bias (PB%)**	**Variance**	**MSE**	**Bias (PB%)**	**Variance**	**MSE**	**Mean**	**Variance**	$P(4 \leq S_{25} \leq 6)$
1	-0.210 (-10.4)	0.531	0.575	-0.345 (-17.1)	145.357	145.476	3.838	0.830	0.636
2	-0.179 (-8.9)	0.491	0.523	-0.142 (-7.0)	40.772	40.792	4.523	0.888	0.856
3	-0.153 (-7.6)	0.451	0.474	-0.041 (-2.0)	3.510	3.512	5.473	0.954	0.856
4	-0.049 (-2.4)	0.315	0.317	0.004 (0.2)	23.854	23.854	6.385	1.104	0.566
5	0.175 (8.7)	0.171	0.201	-0.158 (-7.8)	233.677	233.702	7.555	1.568	0.196

[A] Note: PB refers to the bias divided by the true $\text{MTD}_{0.2} = 2.02$ in percent (i.e. percent bias).

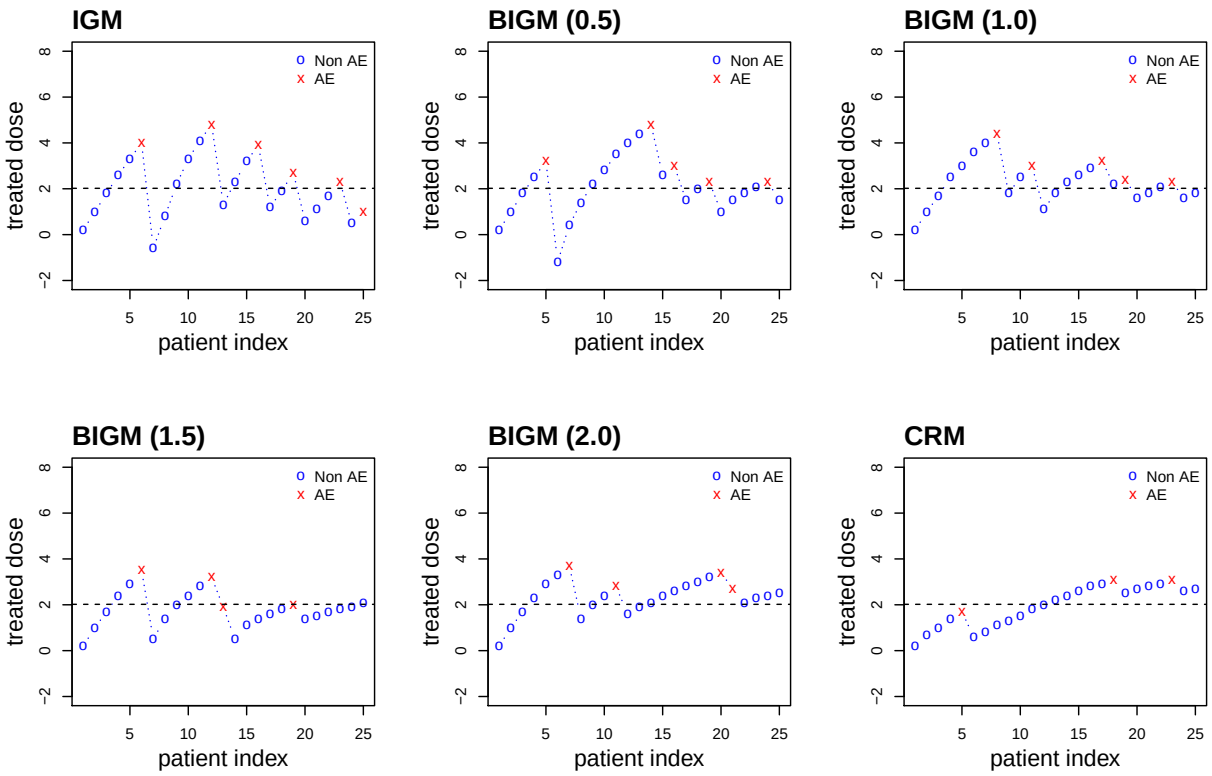


Figure 3.3: Comparing sampling paths of IGM, BIGM and CRM in a small sample

the CRM tends to move slowly through the dosing range. As the trial proceeds, and more information on the true MTD is obtained, the dose allocation behavior of the BIGM begins to resemble that of the CRM. The acceleration of this process increases with increasing λ. In particular, the path of the BIGM with $\lambda = 0.5$ resembles the path of the IGM, and the path of the BIGM with $\lambda = 2.0$ resembles the path in the CRM.

Table 3.1 summarizes simulation results for the IGM, the BIGM with $\lambda = 0.5, 1.0, 2.0$ and the CRM. Specifically, Table 3.1 yields characteristics of the distribution of posterior mean for $\text{MTD}_{0.2}$, the MLE for $\text{MTD}_{0.2}$, and of $S_{25} \equiv \sum_{i=1}^{25} Y_i$ (the total number of observed AEs at trial completion after $N = 25$ subjects have been treated). We first focus on the distribution of the posterior estimate. In general, as λ increases, absolute bias in the posterior mean tends to decrease while variance tends to increase for a given prior. Focusing on Prior 3 (the

prior yielding an estimate of $\text{MTD}_{0.2}$ that is closest to the truth), the resulting mean square errors (MSEs) in the BIGM are generally smaller than the resulting MSEs in the CRM and are close to the IGM for small values of λ. In that prior, the resulting MSEs are 0.405, 0.396, 0.409, 0.425 and 0.474 in the IGM, the BIGM with $\lambda = 0.5, 1.0, 2.0$ and the CRM, respectively. In Prior 1 (the most conservative prior yielding a relatively low estimate of $\text{MTD}_{0.2}$), the respective resulting MSEs are 0.723, 0.641, 0.623, 0.623 and 0.575. The wider sampling scheme reduces variability in posterior estimation but increases the magnitude of underestimation. As a result, the IGM has the maximum MSE despite having minimum variance. In Prior 5 (prior estimate of $\text{MTD}_{0.2}$ is relatively high), the respective results are 0.225, 0.250, 0.250, 0.232, and 0.201. Ironically, the smallest MSE is achieved in the CRM by failure of preventing early trial participants from severe underdosing and overdosing. Although the simulated MSEs of the BIGM are not always smaller than the MSEs of the IGM and the CRM (e.g. Prior 5), the BIGM achieves a reasonable compromise between the IGM and the CRM in most cases with respect to the trade-off between bias and variance in posterior estimation of $\text{MTD}_{0.2}$.

The operating characteristic of each design is also well reflected through the resulting distribution of the MLE for $\text{MTD}_{0.2}$. In particular, the IGM results in a more stable distribution for the MLE when compared to the CRM. The narrow sampling scheme of the CRM occasionally results in an extreme MLE particularly when observed S_{25} is too small or too large. In theory, the expectation of the MLE for $\text{MTD}_{0.2}$ is unbounded because $P(S_{25} = 0) > 0$ in small samples. Observing $S_{25} = 0$ tended to be rare in the presented simulation studies, but the variance of the MLE under the CRM was still observed to be large, particularly when using a prior where the estimate of $\text{MTD}_{0.2}$ was far away from the truth (e.g. variances of the MLE were estimated to be 145.357 under Prior 1 and 233.677 under Prior 5). What is observed is that the resulting distribution of the MLE is very sensitive to prior mis-specification when the CRM is used, but the variance of the MLE gradually decreases when the BIGM is used, particularly when λ approaches zero (converging toward the IGM). By eliminating

prior information in the final estimate of the MTD, the bias of the MLE is close to zero in most cases. As such, from Priors 1 to 5 the resulting MSEs are 0.569, 0.568, 0.570, 0.570 and 0.700 in the IGM, 1.095, 0.751, 0.684, 0.501, 0.884 in the BIGM with $\lambda = 0.5$, and 145.476, 40.792, 3.512, 23.854 and 233.702 in the CRM.

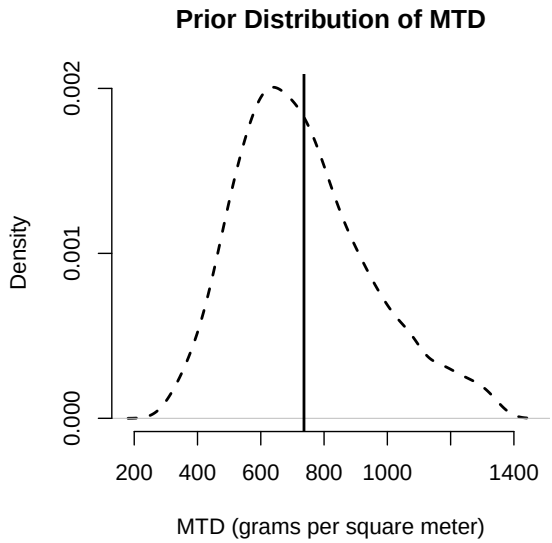
Finally we turn our focus to the resulting distribution of S_{25}. For $N = 25$ and $\gamma = 0.2$, a desired distribution of S_{25} should have a single mode near $N \times \gamma = 5$ with a sharp peak. In the IGM, the BIGM with $\lambda = 0.5, 1.0, 2.0$ and the CRM, the respective probability of observing 4 to 6 AEs at the completion of a given trial was estimated to be 1.000, 0.951, 0.866, 0.797 and 0.636 for Prior 1, 0.980, 0.948, 0.913, 0.876, and 0.856 for Prior 3, and 0.513, 0.434, 0.383, 0.320 and 0.196 for Prior 5. In each case, the narrow sampling scheme of the CRM results in the lowest probability of observing a total number of AEs in the target range. Conversely, the wide sampling scheme of the IGM results in the highest observed probability in each case and the lowest variability in the total number of observed AEs. As intended the BIGM serves as a compromise between the two approaches. Comparing with the resulting distributions of the posterior estimate of $\text{MTD}_{0.2}$, the total number of observed AEs landing in the target range is partially attributable to increased precision in the estimate of $\text{MTD}_{0.2}$.

To summarize the simulation results, the IGM tends to yield more precise estimates of the MTD and hence smaller variability in the total number of observed AEs across simulated trials. In contrast, the CRM tends to yield lower bias in the estimated MTD though with lower precision and high variability in the total number of observed AEs. The proposed BIGM performs as intended, seeking to balance the positive operating characteristics of each procedure. The result is a class of dosing algorithms that yields relatively low MSE for the estimated MTD, low variability in the total number of observed AEs, and high probability of observing a total number of AEs in the anticipated target range. , The choice of λ, and hence the level of compromise between the IGM and CRM, can be made relative to the clinical setting being investigated. Specifically, in cases where less severe AEs are anticipated, values

of λ closer to 0 (approximating the behavior of the IGM) may be preferred. In cases were severe AEs are anticipated, large values of λ (approximating the behavior of the CRM) may be preferred along with the possible addition of overdose control procedures.

3.4 Design of a HIPEC Dosing Trial

Figure 3.4: Approximate prior distribution of the MTD in the HIPEC trial



Hyperthermic intraperitoneal chemotherapy (HIPEC) is used to treat abdominal cancers including ovarian, stomach, and peritoneal cancers. It has been hypothesized that chemotherapy becomes more effective without increasing the strength of the chemotherapy when active chemicals are heated. The treatment has been used based on clinicians' common knowledge without a formal determination of the MTD. Here we consider use of the proposed the BIGM for estimating the MTD of HIPEC in a proposed clinical study. Specifically, in the HIPEC dosing study the target risk level is $\gamma = 0.35$ and the goal is to determine $\text{MTD}_{0.35}$ based upon a total trial sample size of $N = 30$. Six experimental doses are to be considered with values of 450, 600, 750, 900, 1050, and 1200 g/m^2. Hence the goal is to determine the MTD

corresponding to the closest dose level of the estimated $\text{MTD}_{0.35}$. Unlike typical Phase I clinical trials, an informative prior is available because the HIPEC treatment has been practiced in the past. For a conditional mean prior, the research team selected two doses at $x_{-1} = \log(450)$ and $x_0 = \log(1200)$ with $(a_{-1}, b_{-1}) = (1.38, 9.63)$ and $(a_0, b_0) = (3.06, 1.16)$. When compared to the simulation setting defined in Section 3.1, the HIPEC trial is different in that the action space is discrete with six possible experimental doses, the target risk level is greater than $\gamma = 0.2$, and the prior specification is fairly strong. Figure 3.4 presents the approximate induced prior distribution of $\text{MTD}_{0.35}$ with the vertical line indicating the prior mean of 760 g/m^2.

We compare the operating characteristics of the IGM, the BIGM with various values of λ and the CRM with respect to the distribution of estimated MTD, of dose allocation, and of $S_{30} = \sum_{i=1}^{30} Y_i$. We consider six scenarios where the true $\text{MTD}_{0.35}$ is located at the i^{th} experimental dose in the i^{th} scenario. In Scenario 1, the true probabilities of an AE at the six experimental doses are (0.35, 0.45, 0.53, 0.60, 0.65, 0.69). In Scenarios 2 to 6, the respective probabilities are (0.15, 0.35, 0.56, 0.72, 0.83, 0.89), (0.15, 0.25, 0.35, 0.45, 0.53, 0.60), (0.15, 0.22, 0.29, 0.35, 0.41, 0.46), (0.15, 0.20, 0.26, 0.31, 0.35, 0.39), and (0.15, 0.20, 0.24, 0.28, 0.32, 0.35). These scenarios are generated under the two-parameter logistic model.

Table 3.2 summarizes simulation results considered in the design of the HIPEC trial. We first focus on the first two scenarios when the prior guess, $E(\text{MTD}_{0.35}) \approx 750$ g/m^2, is above the true MTD. In Scenario 1 (MTD = 450), the resulting probability of correct posterior guess for MTD is 0.451, 0.445, 0.427, 0.426 and 0.415 in the IGM, the BIGM with $\lambda = 0.5, 1.0, 2.0$ and the CRM, respectively. In Scenario 2 (MTD = 600), the respective result is 0.824, 0.798, 0.788, 0.771, and 0.780. The IGM provides the most accurate final decision at the cost of increased variance in dose allocation. The BIGM makes the compromise between the trade-off. In Scenario 1, when the experimental drug is relative toxic, $E(S_{30}) = 13.0$ in the IGM and $E(S_{30}) = 13.6$ in the CRM, and the BIGM gradually decreased the average

as λ increases. When the true MTD is at the low end of the action space, $Var(S_{30})$ is the maximum in the IGM as opposed to the case of continuous action space, while the variance is smaller in the BIGM than both IGM and CRM. In Scenario 2, $E(S_{30}) = 11.5$ and $Var(S_{30}) = 0.6$ in the IGM, and $E(S_{30}) = 12.2$ and $Var(S_{30}) = 2.0$ in the CRM. In this scenario, the BIGM shows the monotonic trend in both $E(S_{30})$ and $Var(S_{30})$ with respect to the tuning parameter λ. This trend is similar to the case of continuous dosing schedule. In these two scenarios, the BIGM provides a reasonable compromise with respect to the posterior estimation, dose allocation and the number of AEs.

We now focus on Scenarios 3, 4 and 5, when the prior guess is close to the true MTD or below. In these cases, the information gain procedure is not useful. In Scenario 3 (MTD = 750), all designs yield the probability of correct guess between 0.514 and 0.529, whereas the probability of dose allocation at the maximum dose 1200 g/m^2 is 0.114 in the IGM and 0.015 in the CRM. The degree of under- and overdosing is reduced in the BIGM by increasing λ. In Scenario 3, $E(S_{30}) = 10.8$ in the CRM which is about 0.3 greater than the results in the IGM and BIGM. On the other hand, $Var(S_{30}) = 2.2$ in the CRM is substantially greater than $Var(S_{30}) = 0.9$ in the IGM, while the variance in the BIGM gradually decreases as λ decreases toward zero. At the cost of high variability in dose allocation, the IGM obtains a stable behavior of S_{30}. Similar trends in the posterior estimation, dose allocation and S_{30} are found in Scenarios 4 and 5.

In Scenario 6 (MTD = 1200), when the prior guess is severely lower than the true MTD, a deviation by one dose level from the true MTD appears to be tolerable because the true dose-response curve rises gently. In this scenario, the probability of an AE at doses 1050 and 1200 g/m^2 is 0.32 and 0.35, respectively. The resulting probability of the posterior guess at one of the two doses is 0.599, 0.578, 0.572, 0.573 and 0.571 in the IGM, the BIGM with $\lambda = 0.5, 1.0, 2.0$ and the CRM, respectively. The respective probability of dose allocation at the two doses is 0.501, 0.447, 0.422, 0.398 and 0.416. In this scenario, the BIGM tends to

reduce the degree of under estimation at the end of a trial and of underdosing during a trial. All five designs show similar average and variance of S_{30}.

When we alter the prior specification more conservatively so that $E(\text{MTD}_{0.35}) \approx 550$ g/m^2, the BIGM shows the balancing properties in Scenarios 3, 4, 5 and 6, but not as clearly in Scenarios 1 and 2 because the action space is bounded at the low end. It implies that the BIGM (and the IGM as a special case) requires a sufficient room for dose allocation in order to show its intended properties.

3.5 Discussion

In this section, we provide a tractable interpretation of the loss function utilized by the IGM based on its mathematical decomposition. Equation (3.2) reveals that the dose allocation rule in the IGM indeed balances individual-level ethics by a term in the numerator and population-level ethics by a term in the denominator, but the balancing process is rather slow as illustrated in Figure 3.1. Therefore, the loss function derived from the asymptotic variance of the MLE for MTD_γ may not be suitable in small samples, particularly when serious AEs are to be expected. As, we sought to accelerate the compromising procedure by gradually weighing individual-level ethics heavier as a trial proceeds with additional acceleration after each observed AE. To this end, we have proposed an allocation procedure that initiates dosing with an IGM-like sampling scheme to increase precision in the MTD estimate, then make a smooth transition to a CRM-like sampling scheme to reduce potential severe over- and underdosing.

Our proposed procedure provides freedom in the choice of the acceleration tuning parameter, since the pace of the transition should depend upon the clinical setting being considered. Simulation results illustrate the operating characteristics of the proposed BIGM with respect

to balancing the distribution of the MTD estimate and the distribution of the total number of observed AEs at trial completion. The choice of tuning parameter, λ, is non-trivial and it should be trial-specific depending upon severity of potential AEs and anticipated consequences from treatment failure. Based on the results presented here and extended simulation studies, $0.5 \leq \lambda \leq 2.0$ in the proposed BIGM seems to provide a reasonable balance between the two conflicting perspectives in the CRM and the IGM. Before initiating a trial, one possible and recommended approach is to select λ after simulating several scenarios based on all available prior knowledge and various sensitivity parameterizations.

Recently, Oron and Hoff [47] discussed that the number of cohorts treated at the dose level closest to MTD_γ is highly variable in the CRM. They used the terms "long-memory" to describe designs that allocate new patients at an estimated MTD_γ using all observations and "short-memory" to describe designs that use only recent observations for a new decision. The sampling schemes in short-memory designs are wider than in long-memory designs, and they showed short-memory designs tend to be more robust in small-sample dose-finding studies. They also pointed out that the CRM tends to settle early on a specific dose level which is not necessarily the MTD in small samples. In addition, the CRM is more sensitive to a prior. From a practical standpoint, we are concerned with variability in dose allocation rather than the amount of information influencing a new decision. However, from a statistical point of view, eliminating a portion of data to gain robustness may not be appealing. In contrast to short-memory designs, the proposed BIGM is able to adjust variability in dose allocation based on all observed data.

Table 3.2: Simulation results (prospective HIPEC trial)

SCENARIO 1 (MTD = 450)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.451	0.455	0.086	0.007	0.001	0.000	0.534	0.210	0.126	0.073	0.022	0.035	13.034	3.514
BIGM (0.5)	0.445	0.465	0.083	0.007	0.000	0.000	0.440	0.348	0.137	0.053	0.012	0.011	13.016	2.875
BIGM (1.0)	0.427	0.482	0.083	0.006	0.001	0.000	0.387	0.405	0.138	0.053	0.010	0.006	13.116	2.918
BIGM (2.0)	0.426	0.481	0.085	0.007	0.000	0.000	0.332	0.455	0.151	0.049	0.008	0.003	13.229	3.138
CRM	0.415	0.485	0.093	0.006	0.000	0.000	0.240	0.501	0.195	0.054	0.008	0.001	13.586	3.284
SCENARIO 2 (MTD = 600)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.033	0.824	0.143	0.000	0.000	0.000	0.340	0.327	0.205	0.097	0.022	0.008	11.544	0.592
BIGM (0.5)	0.038	0.798	0.163	0.001	0.000	0.000	0.199	0.518	0.216	0.053	0.010	0.003	11.662	1.070
BIGM (1.0)	0.047	0.788	0.165	0.001	0.000	0.000	0.156	0.573	0.215	0.048	0.005	0.001	11.755	1.463
BIGM (2.0)	0.055	0.771	0.172	0.001	0.000	0.000	0.113	0.620	0.222	0.041	0.004	0.001	11.879	1.813
CRM	0.077	0.780	0.141	0.001	0.000	0.000	0.066	0.656	0.242	0.034	0.002	0.000	12.224	1.978
SCENARIO 3 (MTD = 750)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.005	0.221	0.529	0.209	0.030	0.005	0.173	0.256	0.217	0.162	0.079	0.114	10.455	0.927
BIGM (0.5)	0.005	0.214	0.524	0.214	0.036	0.006	0.081	0.277	0.334	0.188	0.063	0.056	10.461	1.430
BIGM (1.0)	0.005	0.217	0.528	0.211	0.036	0.005	0.060	0.275	0.376	0.198	0.056	0.035	10.475	1.681
BIGM (2.0)	0.004	0.220	0.514	0.212	0.042	0.008	0.040	0.273	0.411	0.202	0.051	0.024	10.466	2.071
CRM	0.005	0.219	0.525	0.205	0.040	0.006	0.008	0.240	0.454	0.227	0.057	0.015	10.759	2.236
SCENARIO 4 (MTD = 900)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.002	0.086	0.336	0.325	0.152	0.099	0.124	0.193	0.178	0.163	0.088	0.254	9.674	1.609
BIGM (0.5)	0.002	0.086	0.320	0.335	0.157	0.100	0.059	0.174	0.265	0.229	0.108	0.165	9.650	1.939
BIGM (1.0)	0.002	0.078	0.334	0.336	0.158	0.092	0.042	0.162	0.299	0.259	0.114	0.123	9.622	2.119
BIGM (2.0)	0.001	0.083	0.332	0.330	0.156	0.098	0.027	0.159	0.327	0.279	0.117	0.091	9.582	2.340
CRM	0.001	0.079	0.335	0.329	0.159	0.098	0.004	0.126	0.347	0.305	0.134	0.084	9.803	2.530
SCENARIO 5 (MTD = 1050)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.002	0.046	0.208	0.283	0.206	0.255	0.111	0.151	0.147	0.150	0.083	0.358	8.998	2.665
BIGM (0.5)	0.001	0.041	0.210	0.289	0.202	0.256	0.055	0.129	0.215	0.220	0.119	0.262	8.964	2.651
BIGM (1.0)	0.001	0.043	0.216	0.294	0.200	0.246	0.039	0.120	0.238	0.251	0.139	0.213	8.956	2.623
BIGM (2.0)	0.001	0.041	0.214	0.302	0.208	0.234	0.025	0.115	0.261	0.278	0.151	0.170	8.900	2.686
CRM	0.001	0.044	0.210	0.302	0.216	0.227	0.002	0.089	0.266	0.302	0.175	0.164	9.141	2.893
SCENARIO 6 (MTD = 1200)														
	Posterior Estimation of MTD						**Dose Allocation**						**Number of AEs**	
Method	**450**	**600**	**750**	**900**	**1050**	**1200**	**450**	**600**	**750**	**900**	**1050**	**1200**	**Mean**	**Variance**
IGM	0.001	0.026	0.140	0.234	0.204	0.395	0.110	0.126	0.125	0.138	0.076	0.425	8.425	3.315
BIGM (0.5)	0.001	0.028	0.153	0.240	0.206	0.372	0.057	0.108	0.183	0.205	0.117	0.330	8.481	3.216
BIGM (1.0)	0.001	0.031	0.154	0.241	0.208	0.364	0.042	0.104	0.200	0.233	0.142	0.280	8.462	3.124
BIGM (2.0)	0.000	0.028	0.155	0.244	0.220	0.353	0.025	0.097	0.218	0.262	0.161	0.237	8.414	3.036
CRM	0.001	0.030	0.150	0.249	0.211	0.360	0.002	0.074	0.229	0.280	0.180	0.235	8.608	3.196

Chapter 4

Estimation of Benchmark Dose in the Presence or Absence of Hormesis

In Chapters 2 and 3, we focused on human-based experiments where trial participants are actual patients to be treated. We turn our attention to dose-response modeling for animal-based experiments in Chapters 4, 5 and 6. In Chapter 4, the parameter of interest is termed a benchmark dose in cancer risk assessments, and this quantity is analogous to a maximum tolerable dose in early phase clinical trials. As opposed to Phase I clinical trials, we assume that a background effect of a carcinogen (cancer-causing agent) exists. In other words, animals (or humans) may naturally develop cancers in the absence of the cancer-causing agent being studied. Under current practice of the Environmental Protection Agency, a benchmark dose is defined as the dose associated with a fixed increase of risk relative to the background risk, and an estimated benchmark dose serves as a starting point for downward extrapolation. There are broad two classes of dose-response models, monotonicity and non-monotonicity. As an estimated benchmark dose is sensitive to an assumed dose-response model, we take both classes of assumptions as possibilities for robust estimation.

4.1 Introduction

Mathematical modeling of binomial responses at fixed experimental doses is commonly performed by the Environment Protection Agency (EPA) using dose-response models built in the EPA Benchmark Dose Software (BMDS). For a given model, the parameter of interest is a benchmark dose (BMD) that is a function of regression parameters. In toxicity assessments, a BMD is the dose corresponding to a specified relative increase in the risk of a toxic event when compared to the background risk. The specified increase is referred to as a benchmark risk (BMR) and is often fixed at 0.01, 0.05, or 0.10 in practice.

For estimation of the BMD it is common for investigators to assume a monotonic dose-response relationship [26]. As an alternative to the monotonic assumption, the hormesis theory has gained popularity in recent decades. Hormesis refers to a "beneficial" effect of an agent at low doses and a "harmful" effect at high doses. The hormesis theory is based upon the assumption of stimulatory effects caused by low level exposures to harmful agents. Such dose-response relationships can be visually described as a J-shaped curve at low doses. Calabrese and Baldwin quantified evidence of hormesis and categorized approximately 350 studies into high, moderate and low hormetic effects [11]. The method was based upon five criteria including the number of experimental doses below a hormetic zone (a set of doses corresponding to a lower risk than the background risk), estimation of the hormetic zone, statistical significance of the stimulatory responses, the magnitude of the stimulatory responses, and the replicability of the results. Most toxicology studies do not have a sufficient number of unique doses for regression; a parametric model of the underlying dose-response relationship is nearly always employed. Following this work, multiple studies have elevated the importance of hormesis in toxicology by reporting empirical evidence for hormesis using a series of ad hoc methodologies [10, 14, 15, 19]. In particular, Calabrese and Cook discussed various advantages to parameter estimation if the possibility of a hormetic function was accepted as the default dose-response relationship in carcinogen risk assessments [15]. They

point out that one possible advantage is the protection of both normal and high risk groups by estimating a BMD where the corresponding risk is lower than the background for both groups. They further pointed out that ignoring hormesis leads to a less than ideal regulatory policy. While valid points regarding the costs of ignoring the potential for hormesis have been raised, other authors have pointed out limitations of the ad hoc estimation approaches typically used in previous studies and the lack of validated hypothesis testing procedures for hormetic relationships in existing databases [42]. It has been argued that the current practice of using a lower confidence limit of BMD (BMDL) or an averaged BMDL from various monotonic models is conservative and health-protective, particularly when evidence for hormesis is not definitive [26, 61].

In small-sample studies, a parametric approach requiring a model-based assumption about the underlying dose-response relationship is often implemented for inference regarding a BMD in order overcome sparseness in the data. To reduce the impact of model misspecification, we account for model uncertainty via Bayesian model averaging (BMA) [50]. That is, we do not assume a single dose-response class, but rather take the perspective of BMA by considering both monotonic and hormetic effects as possibilities at low doses. In the BMA approach, a model that is better supported by empirical evidence will yield a higher contribution to the posterior inference of the parameter of interest.

Implementation of model averaging techniques has been extensively discussed in the community of risk assessment based on various information criteria including Akaike information criterion, Bayesian information criterion, and Kullback information criterion [9, 65, 4, 41]. Shao and Small illustrated the application of BMA using Gibbs sampling for the estimation of the BMD. They further determined optimal future experimental doses using logistic and quantal-linear models [56]. Despite the growing popularity of BMA approaches in the risk assessment community, to date these methods have only been described for use in averaging over multiple models that all assume a monotonic dose-response relationship. In this case

the resulting BMD estimate is obtained by averaging over a class of models that give zero prior probability to the possibility of a hormetic effect at low doses.

There are many flexible dose-response models, outside the EPA BMDS, that are suitable for modeling hormesis [40, 33]. Recently, the impact of assuming monotonicity under existence of hormesis has been investigated in a frequentist framework [8]. Bogen studied the impact of misspecification by generating a hormetic curve and fitting several monotonic models using the BMDS. He used a multistage model with a quadratic term to generate the hormetic curve, which he referred to as a generic hockey-stick (GHS) model. He reported that the GHS model performed as well or better than the monotonic models built in BMDS in estimation of the BMD.

In this chapter we consider estimation of the BMD using BMA under two broad dose-response model classes: monotonicity and hormesis. Among many dose-response models in the BMDS, we narrow the scope of our discussion to a multistage model using a Bayesian framework. A general description of the (monotonic) linearized multistage model is given by Crump [20]. Armitage and Doll discussed the two-stage theory of carcinogenesis [2], and Armitage later discussed that models with several stages (e.g. five to seven) that have often been regarded as implausible in the absence of direct biological evidence about a succession of stages [1]. Armitage further mentioned that one can either assume more than two stages or regard some intermediate stages as being fictional shorthand for a single phase. From a statistical perspective, a model with more than three parameters may suffer from over-fitting when there are four to six experimental dose groups (including a control group) as commonly encountered in past experiments for cancer risk assessment [71].

To incorporate a hormetic class of models in the BMA approach we consider a two-stage model (i.e. up to a quadratic term with three parameters) that allows hormetic effects at low doses rather than the usual assumption of monotonicity. The proposed parameterization can be also thought of as weighting two mutually exclusive parameter spaces, where the

posterior weights are dictated by changes in the dimension and support of the underlying model parameters.

This chapter is structured as follows. In Section 4.2, we introduce the statistical methodology including our proposed extension to a two-stage hormetic class of models and the BMA algorithm used to average over classes models derived from assumption of monotonicity and hormesis. In Section 4.3, we present the results of simulation studies designed to investigate operating characteristics of the proposed two-stage model and robustness of the BMA method under a variety of monotonic and hormetic dose-response settings. In Section 4.4, we apply the proposed methodology to two datasets previously reported by Kociba et al and the NTP study that consider the carcinogenic toxicity of 2,3,7,8-Tetrachlorodibenzo-p- Dioxin (TCDD) [38, 43]. Section 4.5 concludes with a discussion on the use of less restrictive modeling assumptions in risk assessment studies and avenues for future research.

4.2 Methods

4.2.1 Linearized Multi-Stage Model for Potential Hormesis

Let $\pi_x(\vec{\beta})$ denote the probability of an event associated with dose $x \geq 0$, where $\vec{\beta} = (\beta_0, \ldots, \beta_M)^T$ denotes a vector of model parameters. Assuming $Y \sim \text{Bernoulli}\left[\pi_x(\vec{\beta})\right]$, a monotonic dose-response relationship under a linearized multistage model is given by

$$\pi_x(\vec{\beta}) = P(Y = 1 \mid d) = 1 - e^{-\sum_{m=0}^{M} \beta_m x^m}$$

for $\beta_m \geq 0$. The multistage model is used by the EPA for cancer risk assessments [26], typically with $M = 2$. The background probability at $x = 0$ is given by $\pi_0(\beta_0) = 1 - e^{-\beta_0}$,

therefore $\beta_0 = -\log(1-\pi_0)$. By taking the derivative of $\pi_x(\vec{\beta})$ with respect to x,

$$\frac{\partial \pi_x(\vec{\beta})}{\partial x} = e^{-\sum_{m=0}^{2}\beta_m x^m}(\beta_1 + 2\beta_2 x)\,. \tag{4.1}$$

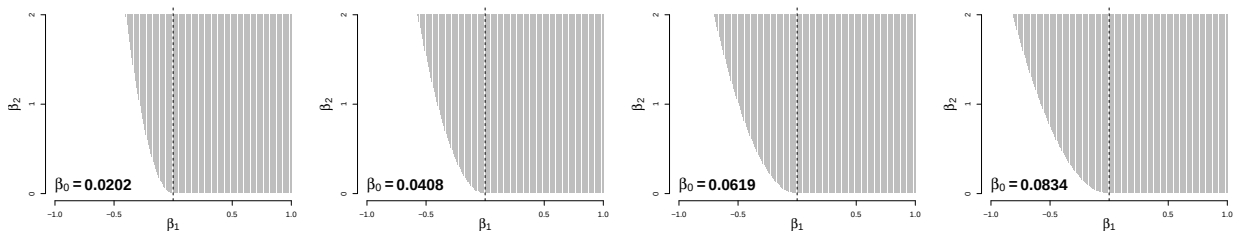
Since the above derivative approaches $\beta_1 e^{-\beta_0}$ as $x \to 0$, the second parameter $\beta_1 \geq 0$ guarantees monotonicity. For $x > 0$ the rate of increasing toxicity is determined by β_2.

In order to model a hormetic effect at low doses (i.e. beneficial effect at low doses), we assume $\beta_0 > 0$ (i.e. $\pi_0 > 0$), $\beta_1 < 0$, and $\beta_2 > 0$. To allow for both monotonicity and hormesis in the model, we consider $-\infty < \beta_1 < \infty$ with constraint $0 < \pi_x(\vec{\beta}) < 1$ for all $x > 0$. If a true dose-response relationship is hormetic, the minimum probability of an event occurs at $x = -\beta_1/(2\beta_2) > 0$ from Equation (4.1), and we let δ denote that most beneficial dose. It can be verified that the second derivative $\partial^2 \pi_x(\vec{\beta})/\partial d^2$ evaluated at $\delta = -\beta_1/(2\beta_2)$ is positive, hence we have a J-shape curve at the hormetic zone. In order to preserve $0 < \pi_x(\vec{\beta}) < 1$, we guarantee that a linear predictor, $\sum_{m=0}^{2}\beta_m x^m > 0$, for all $x > 0$. Finally, since we need $\min_{x>0}\pi_x(\vec{\beta}) = \pi_\delta(\vec{\beta}) > 0$, we require $\sum_{m=0}^{2}\beta_m \delta^m > 0$. The parameter space for $\vec{\beta} = (\beta_0, \beta_1, \beta_2)^T$ can then be partitioned as $\Omega = \Omega_M \cup \Omega_H$, where $\Omega_M = (0,\infty)\times(0,\infty)\times(0,\infty)$ and

$$\Omega_H = \left\{ \vec{\beta} \in (0,\infty)\times(-\infty,0)\times(0,\infty) : \beta_2 > \frac{\beta_1^2}{4\beta_0} \right\}. \tag{4.2}$$

In other words, we increase the flexibility of the two-stage model by expanding the original parameter space Ω_M. In this parameterization, we have $\Omega(\beta_0^{(1)}) \supset \Omega(\beta_0^{(2)})$ whenever $\beta_0^{(1)} > \beta_0^{(2)}$ (i.e. $\pi_0^{(1)} > \pi_0^{(2)}$). Figure 4.1 shows the two-dimensional space of (β_1, β_2) for $\pi_0 = 0.02$, 0.04, 0.06, and 0.08. When $\vec{\beta} \in \Omega_H$ the two-stage model suffers from lack of flexibility because of its intrinsic local symmetry about $x = \delta$. We will further discuss this limitation in later sections.

Figure 4.1: Parameter space under hormesis



4.2.2 Benchmark Dose under the Two-Stage Model

The EPA defines the BMD for a given BMR as the dose that satisfies

$$\frac{\pi_{\text{BMD}} - \pi_0}{1 - \pi_0} = \text{BMR}\,, \tag{4.3}$$

or equivalently $\pi_{\text{BMD}} = \pi_0 + (1 - \pi_0)\,\text{BMR}$ [26]. Under the two-stage model, the BMD satisfies

$$\text{BMR} = 1 - e^{-(\beta_1 \text{BMD} + \beta_2 \text{BMD}^2)}\,. \tag{4.4}$$

for a given BMR. Using the quadratic equation, the closed form for the BMD is given by

$$\text{BMD} = \frac{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_2 \log(1 - \text{BMR})}}{2\beta_2} > 0\,,$$

and exactly one BMD exists even under hormesis (the positive root) because the quantity of the square root is always greater than the magnitude of β_1.

4.2.3 Likelihood and Prior Specification

Consider I experimental doses $0 \leq x_1 < \cdots < x_I$. If $x_1 > 0$, the background risk π_0 is estimated from a parametric assumption. Further suppose $Y_i \sim \text{Binomial}(n_i, \pi_{x_i})$ arise independently, $i = 1, \ldots, I$. The likelihood function is then given by

$$L(\vec{\beta}) = f(\vec{y} \mid \vec{\beta}) \propto \prod_{i=1}^{I} \left[\pi_{x_i}(\vec{\beta})\right]^{y_i} \left[1 - \pi_{x_i}(\vec{\beta})\right]^{n_i - y_i} ,$$

where $\vec{y} = (y_1, \ldots, y_I)$ denotes realized Binomial outcomes.

The prior specification for $\vec{\beta}$ can be done in various ways. Since there is a one-to-one relationship between the background effect π_0 and $\beta_0 = -\log(1 - \pi_0)$, we can specify a Beta distribution for π_0 and sebsequently induce the prior distribution for β_0 using Jacobian transformation. Given β_0, we assign a joint conditional prior distribution for (β_1, β_2) so that the joint support of $\vec{\beta}$ satisfies $\Omega = \Omega_M \cup \Omega_H$. Detailed discussion about a prior specification is provided in Appendix A.3 and Gibbs sampling method for numerically approximate the posterior is provided in Appendix A.4.

4.2.4 Connection to Bayesian Model Averaging

Shao and Small used BMA with two models, logistic and quantal-linear, to account for model uncertainty, but both models were restricted to monotonicity [56]. If we treat a two-stage model with $\vec{\beta} \in \Omega_M$ and a two-stage model with $\vec{\beta} \in \Omega_H$ as two dose-response models, posterior inference for a BMD using a two-stage model with the parameter space $\Omega = \Omega_M \cup \Omega_H$ can be regarded as BMA-based inference based on the two models. Alternatively, the proposed parameterization can be also regarded as weighting two mutually exclusive parameter spaces under the same model, where the posterior weights are determined by the plausibility of the disjoint supports of the model parameters given data.

Using BMA, if μ is a parameter of interest whose interpretation is the same under the two models (e.g. μ = BMD), by the law of total probability, the posterior distribution of μ is given by

$$f(\mu \mid \vec{y}) = f(\mu \mid \vec{y}, \beta_1 < 0)\, P(\beta_1 < 0 \mid \vec{y}) + f(\mu \mid \vec{y}, \beta_1 \geq 0)\, P(\beta_1 \geq 0 \mid \vec{y})\,.$$

Note that the parameter space of $\vec{\beta} = (\beta_0, \beta_1, \beta_2)$ can be partitioned into two sets Ω_M and Ω_H by the sign of β_1. A more formal discussion about the proposed parametrization $\Omega = \Omega_M \cup \Omega_H$ in conjunction with BMA is provided in Appendix A.5.

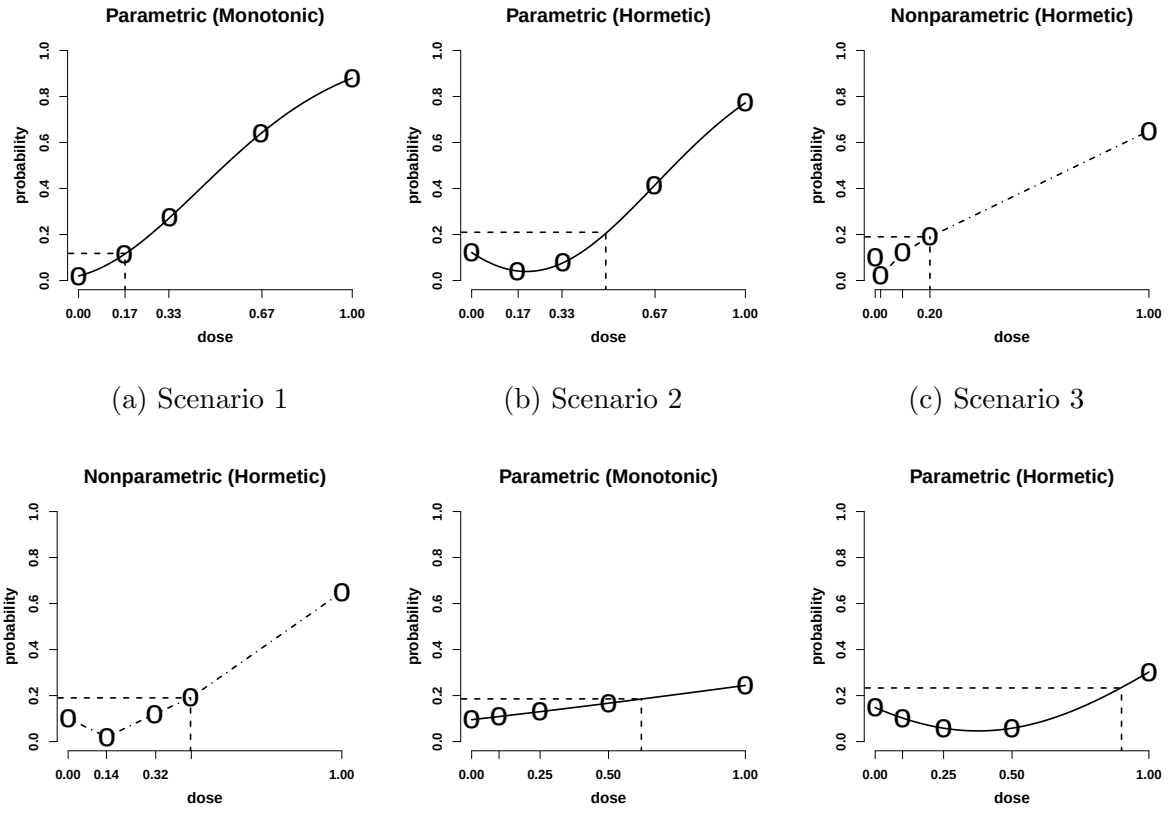
4.3 Simulations

In this section we investigate the impact of assuming monotonicity (Ω_M) when hormesis (Ω_H) is present and visa versa under the two-stage model in the context of BMD estimation (for BMR=0.10). We also examine the robustness of the BMA method under the increased flexibility of the two-stage model with $\Omega = \Omega_M \cup \Omega_H$. We further consider the impact of model misspecification when the true dose-response relationship lies outside the two-stage model parameterization.

4.3.1 Simulation

In the presented simulation studies we assumed that $N = 150$ observations were available with five experimental doses $(d_1, \ldots, d_5)$ such that the five experimental groups had the same sample size (i.e. $n_i = 30$). We take $x_1 = 0$ in order to estimate the background probability π_0. For a prior specification extensively discussed in Appendix A, we assume truncated normal priors for β_1 and β_2 given β_0, and let $\mu = 0$ and $\sigma^2 = 10^2$ (i.e. the mean and the variance before truncation), setting the bounds such that the parameter space satisfies Equation (4.2).

Figure 4.2: Simulation scenarios.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

(d) Scenario 4 (e) Scenario 5 (f) Scenario 6

Table 4.1: Simulation scenarios

Scenario	Experimental Doses	True Probabilities	BMD	Remark
1	$(0.00, 0.17, 1.33, 0.67, 1.00)$	$(0.02, 0.11, 0.27, 0.64, 0.88)$	0.17	Monotonic (parametric) [A]
2	$(0.00, 0.17, 1.33, 0.67, 1.00)$	$(0.12, 0.04, 0.08, 0.41, 0.77)$	0.49	Hormetic (parametric) [A]
3	$(0.00, 0.02, 0.10, 0.20, 1.00)$	$(0.10, 0.02, 0.12, 0.19, 0.65)$	0.20	Hormetic (nonparametric)
4	$(0.00, 0.14, 0.32, 0.45, 1.00)$	$(0.10, 0.02, 0.12, 0.19, 0.65)$	0.45	Hormetic (nonparametric)
5	$(0.00, 0.10, 0.25, 0.50, 1.00)$	$(0.10, 0.11, 0.13, 0.17, 0.24)$	0.62	Monotonic (parametric) [A]
6	$(0.00, 0.10, 0.25, 0.50, 1.00)$	$(0.15, 0.10, 0.06, 0.06, 0.30)$	0.90	Hormetic (parametric) [A]

[A] The dose-response curve was generated from a two-stage model.

To express large prior uncertainty for $\theta = P(\beta_1 < 0)$, we let $\theta \sim \text{Beta}(1.01, 1.01)$ so that 95% of mass was between $\theta = 0.03$ and $\theta = 0.97$. We further set our best guess for π_0 as $\hat{\pi}_0 = 0.10$ (the prior mode) and we assumed 95% certainty that π_0 could not exceed 0.50. Using the bisection method, we could find a_0 and b_0 such that $\hat{\pi}_0 = (a_0 - 1)/(a_0 + b_0 - 1) = 0.10$ and

$$P(\pi_0 < 0.50) = \int_0^{0.50} \frac{\Gamma(a_0 + b_0)}{\Gamma(b_0)\,\Gamma(b_0)}\, \pi_0^{a_0 - 1}(1 - \pi_0)^{b_0 - 1}\, d\pi_0 = 0.95\,.$$

The prior opinion about π_0 was then modeled as $\pi_0 \sim \text{Beta}(a_0 = 1.36, b_0 = 5.12)$.

Based upon the above assumptions, we considered six simulation scenarios with the normalized dose scale $[0, 1]$ (Figure 4.2 and Table 4.1). Scenarios 1 and 2 are under the two-stage models with $(\beta_0, \beta_1, \beta_2)$ = (0.02,0.30,1.80) and (0.13, -0.90, 2.25), respectively. Scenario 1 is a monotonic case, and Scenario 2 is a hormetic case. To investigate the impact of model misspecification, we nonparametrically fixed five probabilities at given experimental doses in Scenarios 3 and 4. Both scenarios are hormetic cases, where Scenario 3 is designed to investigate the impact of severe asymmetry of a hormetic zone while the degree of asymmetry is relatively mild in Scenario 4. As opposed to the first four scenarios in which the probabilities at maximum experimental doses are quite high ($\pi_{x_5} \geq 0.65$), the last two scenarios, Scenarios 5 and 6, were designed so that the probabilities at maximum experimental doses are relatively low ($\pi_{x_5} \leq 0.30$). Under two-stage models, monotonic Scenario 5 used $(\beta_0, \beta_1, \beta_2) = (0.10, 0.15, 0.03)$ and hormetic Scenario 6 used $(\beta_0, \beta_1, \beta_2) = (0.16, -0.60, 0.80)$.

Table 4.2: Simulation results

Parameter Space	Scenario	BMD	Bias	Percent Bias	MSE	Percent $\sqrt{\text{MSE}}$	$P(\text{BMD} \in \text{CI}_{0.95})$
Ω	1	0.17	0.03	20.0	3e-3	31.6	0.95
	2	0.49	-0.06	-11.6	0.01	17.6	0.88
	3	0.20	0.06	27.8	0.01	47.5	0.98
	4	0.45	-0.02	-3.9	3e-3	12.5	0.94
	5	0.62	0.12	19.9	0.06	40.6	0.98
	6	0.90	-0.10	-11.0	0.03	19.1	0.88
Ω_M	1	0.17	-0.01	-4.7	8e-4	15.9	0.98
	2	0.49	-0.22	-44.6	0.05	44.9	0.00
	3	0.20	-2e-3	-0.9	2e-3	23.2	0.98
	4	0.45	-0.12	-27.6	0.02	28.3	0.04
	5	0.62	-0.03	-4.4	0.02	20.7	0.95
	6	0.90	-0.28	-31.6	0.10	34.6	0.57
Ω_H	1	0.17	0.09	50.3	0.01	52.8	0.02
	2	0.49	-0.04	-9.00	0.01	15.3	0.89
	3	0.20	0.19	95.7	0.04	100.9	0.00
	4	0.45	2e-3	0.5	4e-3	13.4	0.97
	5	0.62	0.29	46.1	0.15	61.8	0.68
	6	0.90	-0.05	-5.61	0.02	16.7	0.90

Through the four hormetic scenarios (Scenarios 2, 3, 4, and 6), we investigated the operating characteristics of the proposed parameterization under various lengths of a hormetic zone and locations of a BMD in the unit scale as shown in Figure 4.2.

4.3.2 Simulation Results

Table 4.2 summarizes the results of 10,000 simulations in each of the four scenarios. In Scenario 1, where the true dose-response relationship was generated by the monotonic two-stage model, the expectation of point estimates for the BMD was 0.20, compared to the true BMD of 0.17 (0.03 bias and 20.0% bias), and 95% credible intervals (CIs) for the BMD covered the true value with probability 0.95. Focusing on the conditional results, when we constrain the model fit to $\beta_1 \geq 0$ (i.e. Ω_M) or to $\beta_1 < 0$ (i.e. Ω_H), the performances of the two models were remarkably different. When conditioning on $\beta_1 \geq 0$, the percent bias was -4.7%, the percent $\sqrt{\text{MSE}}$ was 15.9%, and the coverage probability was 0.98. On the other hand, when conditioning on $\beta_1 < 0$, which was the incorrect trend at low doses, the

respective results were 50.3%, 52.8%, and 0.02. We could anticipate severe under-coverage because of the transition from a negative slope to a positive slope at low doses resulting in overestimation of the BMD when the true trend was monotonic. The posterior probability of a hormetic effect, $P(\beta_1 < 0 \mid \vec{y})$, was estimated in this scenario to be 0.47 on average.

Focusing on Scenario 2, where the true dose-response relationship was generated by the hormetic two-stage model, the percent bias and percent $\sqrt{\text{MSE}}$ were -11.6% and 17.6%, respectively, and the coverage probability of 95% CIs was 0.88 marginally. Conditioning on the two parameter spaces separately, the coverage probabilities were 0.00 and 0.89 under Ω_M and Ω_H, respectively. Although the marginal coverage probability of 0.88 was below the nominal 0.95, the results highlight (1) the negative impact of ignoring hormesis when there exists a hormetic zone at low doses and (2) the robustness gained by using the BMA method. On average, $P(\beta_1 < 0 \mid \vec{y}) = 0.92$ in the second scenario.

The results from Scenario 3 illustrate a limitation of the hormetic two-stage model. Recall that the true dose-response relationship was generated with an asymmetric hormetic zone. Since this shape could not be represented via the two-stage model, we should expect some impact of model misspecification. The conditional percent bias and $\sqrt{\text{MSE}}$ were -0.9% and 23.2%, respectively, under Ω_M, while the respective results were 95.7% and 100.9% under Ω_H. The coverage probabilities were 0.98 and 0.00 under Ω_M and under Ω_H, respectively. Without considering the impact of model misspecification, the results were the opposite from our anticipation because the conditional results from Ω_M were more satisfying than the conditional results from Ω_H despite the true dose-response curve being hormetic. By applying the BMA method, however, the percent bias and $\sqrt{\text{MSE}}$ were 27.8% and 47.5% with coverage probability 0.98. An illustration of the results follows. In this scenario, when a posterior draw of $\vec{\beta}$ was in Ω_M (i.e. $\beta_1 \geq 0$), the estimated monotonic dose-response curve often passed through the first two points marked by "O" in the third panel of Figure 4.2. As a consequence, on average, the deviation of an estimated BMD from the true BMD was

not too serious despite the incorrect monotonicity assumption. On the other hand, when a posterior draw for $\vec{\beta}$ was in Ω_H (i.e. $\beta_1 < 0$), the estimated hormetic zone was generally longer than the truth, and such tendency caused the larger positive bias for the BMD. On average, $P(\beta_1 < 0 \mid \vec{y}) = 0.31$ in the third scenario, which is quite small when we consider a hormetic effect truly exists in this scenario.

Revisiting Figure 4.2d, recall that the hormetic zone was approximately symmetric in Scenario 4. When conditioning on Ω_M, we observed -27.6% bias, 28.3% $\sqrt{\text{MSE}}$, and a coverage probability of 0.04. On the other hand, when conditioning on Ω_H, we observed 0.5% bias, 13.4% $\sqrt{\text{MSE}}$, and a coverage probability of 0.97. The BMA with $\Omega = \Omega_M \cup \Omega_H$ yielded -3.9% bias, 12.5% $\sqrt{\text{MSE}}$, and 0.94 coverage probability. On average, $P(\beta_1 < 0 \mid \vec{y})$ was estimated to be 0.78 in this scenario. A take-home message from Scenarios 3 and 4 is that BMA is robust in these simulations not only because it allows for monotonicity or hormesis in the two-stage model, but also because that flexibility allows it to perform better when that model is mis-specified due to asymmetry in the hormetic zone.

The results from Scenario 5 showed the similar trend when compared to the results from Scenario 1. Despite the highest response was relatively lower in the fifth scenario, the magnitude and the direction of bias under Ω_M, Ω_H, and Ω were similar. The results from hormetic Scenario 6 and the results from hormetic Scenario 2 were similar as well. In particular, the smaller absolute bias and MSE and better coverage probability were shown in Ω_H than in Ω_M. In addition, the results in Ω was closer to the results in Ω_H which contained the true parameter values. On average, $P(\beta_1 < 0 \mid \vec{y})$ was 0.46 and 0.80 in Scenarios 5 and 6, respectively. The simulation results from the last two scenarios show the robustness of the BMA with $\Omega = \Omega_M \cup \Omega_H$ when the probability at the maximum experimental dose is quite low. In all scenarios, a larger value of σ does not lead to significantly different results.

4.4 Applications

4.4.1 Hyperplastic Nodule and Hepatocellular Carcinoma

Kociba et al studied toxicity of 2,3,7,8-TCDD including the tumor category of hyperplastic nodules and hepatocellular carcinoma at four fixed doses 0, 1, 10, and 100 ng/kg/day [38]. Hyperplastic nodules are characterized by abnormal growth of thyroid tissue, and hepatocellular carcinoma is one of most common types of liver cancer which is known to be related to viral hepatitis or cirrhosis. The observed proportions of incidence y_i/n_i in Sprague-Dawley rats were $9/86 = 0.105$, $3/50 = 0.060$, $18/50 = 0.360$, and $34/48 = 0.708$ at the four doses, respectively ($N = 234$). The data was recently analyzed by Small and Shao using BMA with two monotonic dose-response models, the logistic and quantal-linear models. In their analysis the posterior means for the BMD (at BMR = 0.10) were 20.93 and 8.08 with 5^{th} percentiles (BMDL) 16.70 and 5.85 under the two models, respectively. They also reported MLEs for the BMD and BMDLs under the eight different models built into the EPA BMDS. The MLEs ranged from 3.226 to 20.667 ng/kg/day, and the BMDLs ranged from 1.244 to 16.722 ng/kg/day under the eight models, which showed model-sensitivity even under the same class of monotonic assumptions.

In the prior elicitation under the two-stage model, we expressed large uncertainty for β_1 and β_2. For β_1 we used the truncated normal distributions $\text{TN}(\mu = 0, \sigma^2 = 10^2, 0, \infty)$ for Ω_M and $\text{TN}(\mu = 0, \sigma^2 = 10^2, -\infty, 0)$ for Ω_H. For $\beta_2 \mid \beta_0, \beta_1$ we used $\text{TN}(\mu = 0, \sigma^2 = 10^2, \beta_1^2/(4\beta_0), \infty)$ for both parameter spaces. For β_0 we transformed $\pi_0 \sim \text{Beta}(a_0, b_0)$ to β_0 using the Jacobian transformation with $a_0 = 1.01$ and $b_0 = 1.01$ to elicit large uncertainty. The same Beta distribution for $\theta = P(\beta_1 < 0)$ was used as in the previously presented simulation studies, $a_\theta = b_\theta = 1.01$, so that $P(\Omega_M) = P(\Omega_H) = 0.5$.

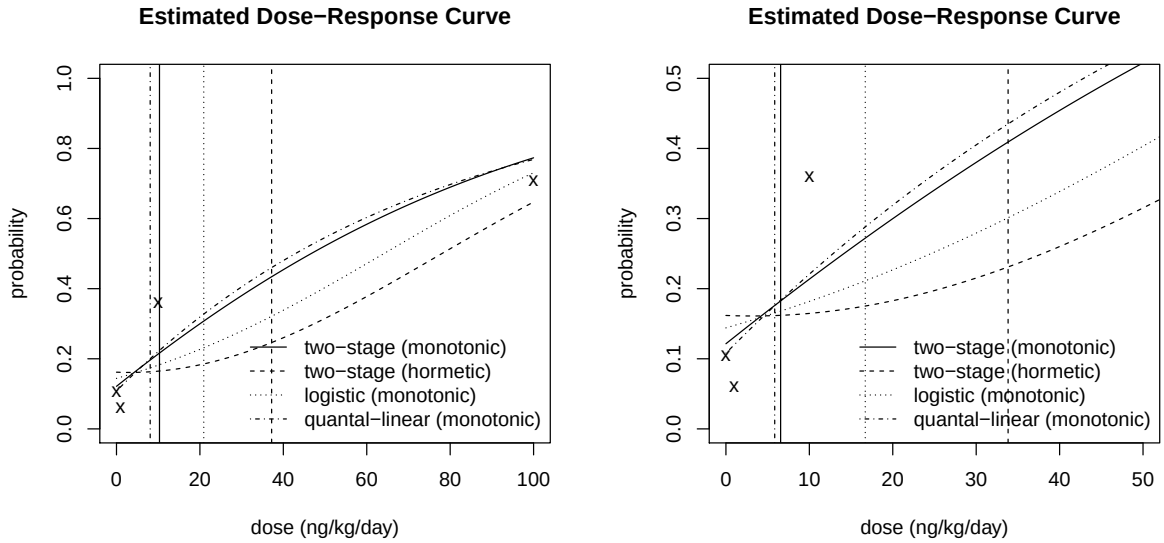
When we restricted the parameter space to Ω_M, the posterior mean of the BMD was 12.83

ng/kg/day with BMDL 7.87 ng/kg/day. For Ω_H the posterior mean of the BMD was 35.28 ng/kg/day with BMDL 27.83 ng/kg/day which were nearly three times greater than the results under the monotonic assumption. For $\Omega = \Omega_M \cup \Omega_H$ utilizing BMA with the equal prior probability on the two classes, the posterior mean of the BMD was 13.78 ng/kg/day with BMDL 7.88 ng/kg/day which were close to the results for Ω_M. In the BMA approach, the posterior mean of β_1 was 0.008 with 95% CI (-0.001, 0.015) which was strong evidence for monotonicity under the model. The small posterior probability $P(\beta_1 < 0 \mid \vec{y}) = P(\Omega_H \mid \vec{y}) = 0.043$ also explained why the BMA-based estimation was highly skewed towards Ω_M (see Table 4.3). In considering sensitivity of the posterior probabilities to the prior specification, it is worth considering the degree of prior weight one would have had to place on a hormetic dose response curve in order to obtain a posterior decision in favor of hormesis. In this case, in order to achieve $P(\Omega_H|\vec{y}) = 0.5$ and 0.9, we would have needed to select (a_θ, b_θ) such that $P(\Omega_H) = 0.957$ and 0.995, respectively. That is, one would have been nearly fully convinced of a hormetic dose response curve prior to experimentation in order to conclude the possibility of hormesis after incorporating the observed data.

If the true value of π_0 was approximately 0.10, then the true BMD (at BMR = 0.10) should satisfy $\pi_{\text{BMD}} \approx 0.10 + (1 - 0.10) \times 0.10 = 0.19$. We could infer that the true BMD might lie between $x_1 = 0$ and $x_3 = 10$ ng/kg/day unless the true dose-response relationship behaved awkwardly. Even if a hormetic zone was truly present, it might be difficult to believe the posterior inference under Ω_H. Instead, the results under Ω_M seem to be more reasonable in this setting. By relying on the robustness of BMA in the estimation of the BMD (see simulation Scenario 3), the BMA-based estimation seems a more reasonable approach when compared to simply conditioning on Ω_H, despite the observed data revealing a decrease in the response probabilities at low doses relative to the background.

Figure 4.3 plots the estimated dose-response curves using the posterior mean of $\vec{\beta}$ under the four different models: two-stage monotonic, two-stage hormetic, logistic, and quantal-

Figure 4.3: Estimated dose-response curves (Kociba et al, 1978)



(a) Whole experimental range

(b) Magnifying low doses

linear models. The estimated curves under the logistic and quantal-linear models were drawn using the posterior means reported by Shao and Small [56]. Figure 4.3a presents the whole experimental range from 0 to 100 ng/kg/day, and Figure 4.3b magnifies the lower-left quadrant of the left panel. The vertical lines in the left panel indicate the posterior means of the BMD conditioning on each model, and the vertical lines in the right panel indicate the corresponding BMDLs. As shown in the figure, the fits under the two-stage monotonic model and the quantal-linear model are very similar because the two models are nested (i.e. $\beta_2 = 0$ in the two-stage). The negative slope in the two-stage hormetic model pulled the estimated BMD and BMDL away from $x = 0$ when compared to the other monotonic models. Although we observed $y_1/n_1 > y_2/n_2 < y_3/n_3$ in the dataset, the two-stage hormetic model could not generally detect the empirical trend during Gibbs sampling due to the extreme asymmetry of the dose response curve about the nadir.

Table 4.3: Estimated BMDs and BMDLs

Tumor Category	**Parameter Space**	$E(\text{BMD} \mid \vec{y})$	BMDL	$P(\beta_1 < 0 \mid \vec{y})$
Hyperplastic nodule, hepatocellular carcinoma	Ω_M	12.83	7.87	0.000
	Ω_H	35.28	27.83	1.000
	Ω	13.78	7.88	0.043
Female rats liver neoplastic nodule	Ω_M	39.97	25.62	0.000
	Ω_H	54.80	42.97	1.000
	Ω	47.52	28.40	0.509

4.4.2 Liver Neoplastic Nodules in Female Rats

Following Kociba et al's study in 1978, the NTP conducted several experiments to investigate the carcinogenesis of 2,3,7,8-TCDD. Here we focus on one particular dataset for the tumor category liver neoplastic nodule [43]. The dataset included Binomial responses in Osborne-Mendel female rats at four experimental doses 0, 1.4, 7.1, and 71 ng/kg/day. The observed proportions of tumor occurrence were $5/75 = 0.067$, $1/49 = 0.020$, $3/50 = 0.060$, and $12/49 = 0.245$, respectively ($N = 223$). Under the logistic and quantal-linear models, by Shao and Small, the posterior means of the BMD were 50.85 and 33.52 ng/kg/day, respectively, and BMDLs were 36.06 and 18.35 ng/kg/day, respectively [56]. Even under the same class of assumptions, the two results greatly differ, particularly the two-fold relative difference in the estimated BMDLs.

We used the same vague prior distributions described in Section 4.4.1 and $P(\Omega_M) = P(\Omega_H) = 0.5$. Under the two-stage monotonic model, the posterior mean of the BMD was 39.97 ng/kg/day with BMDL of 25.62 ng/kg/day. Under the two-stage hormetic model, the posterior mean of the BMD was 54.80 ng/kg/day with BMDL of 42.97 ng/kg/day. Applying the proposed BMA with the equal prior probability on the two classes, the posterior mean of the BMD was 47.52 ng/kg/day with BMDL of 28.40 ng/kg/day with $P(\beta_1 < 0 \mid \vec{y}) = P(\Omega_H \mid \vec{y}) = 0.509$. That is, our posterior belief for the two classes of assumptions, monotonicity and hormesis, has not changed much from the prior belief. In this case, in order to obtain $P(\Omega_H|\vec{y}) = 0.9$, we would have required a prior probability of hormesis of 0.897.

Figure 4.4 plots the estimated dose-response curves under the four models. As in Kociba et al's study, the negative slope in the hormetic model led to the relatively larger estimate of the BMD and BMDL when compared to the other three models. One major difference from the Kociba et al dataset is the less severe asymmetry in the empirical hormetic zone, which partially accounts for the greater posterior probability of Ω_H when compared $P(\Omega_H \mid \vec{y}) = 0.043$ in the Kociba's dataset.

In Figure 4.4a we can see that the four fitted curves passes near the coordinate ($x_4 = 71, y_4/n_4 = 12/49$) due to the high leverage of this dose level. The hormetic model is sensitive to this impact because the leverage point tends to extend the length of an estimated hormetic zone. It is a potential source of overestimation of the BMD under the hormetic model. If the hormetic model truly overestimated the BMD, the potential positive bias could be relieved to some degree by BMA because

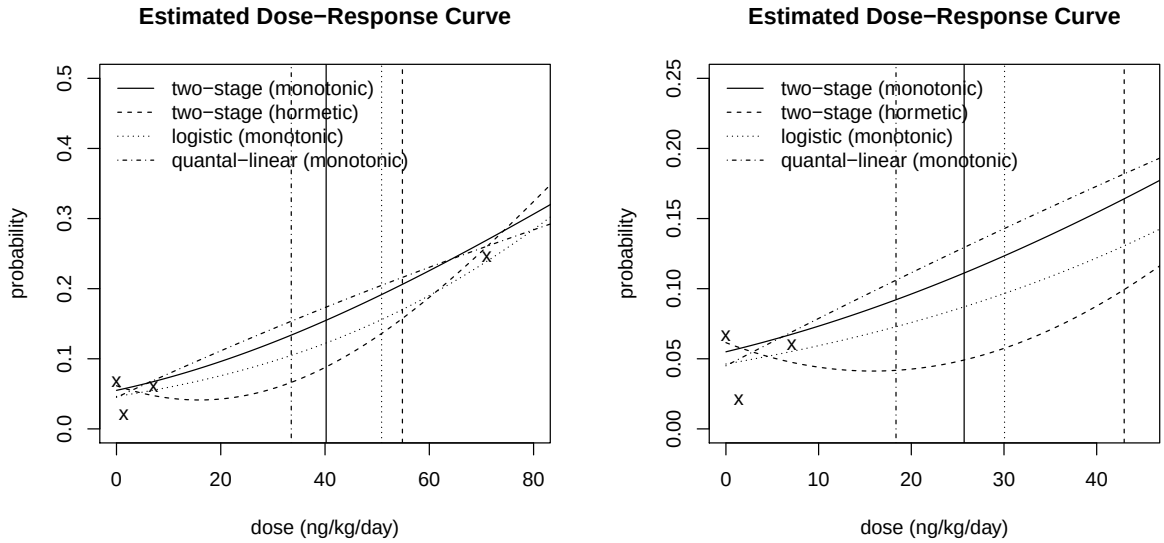
$$E(\text{BMD} \mid \vec{y}, \Omega) = 0.509\, E(\text{BMD} \mid \vec{y}, \Omega_H) + 0.491\, E(\text{BMD} \mid \vec{y}, \Omega_M)\,.$$

Despite $P(\Omega_H \mid \vec{y}) \approx P(\Omega_M \mid \vec{y})$, BMDL = 28.40 under Ω was closer to BMDL = 25.62 under Ω_M alone when compared to an estimated BMDL of 42.97 under Ω_H.

4.5 Discussions

Our discussion has focused on the estimation of the BMD accounting for possible hormetic dose-response relationships. To this end we have extended the nonnegative parameter space in the classic two-stage model. We also illustrated that the two-stage model with $\Omega = \Omega_M \cup \Omega_H$ is equivalent to a BMA approach, while the sign of β_1 determines the class of assumptions. In the presented simulation studies, the BMA approach exhibited robustness in the estimation of the BMD when compared to assuming only one class of assumptions

Figure 4.4: The estimated dose-response curves (NTP, 1982)



(a) Whole experimental range

(b) Magnifying low doses

(either monotonic or hormetic). In particular, undesirable results in the average percent bias in the BMD estimate and coverage probability of 95% CIs could be avoided by the new parameterization. Modification of any generalized linear model to allow a hormetic effect can be done in a similar manner.

Some models in the EPA BMDS, such as multistage models, the logistic model, and the probit model, take an original dose x in the linear predictor. A common pitfall of those models in the estimation of the BMD is the effect of leverage. It can be shown that, under the model with a linear predictor, the second derivative of the log-likelihood with respect to the slope parameter involves the second order of x. As a consequence, the highest experimental dose provides the most information regarding the slope parameter and possibly produces a large bias in the estimate of the BMD at a low BMR. In the two applied examples presented here, we observed that the fitted models tended to chase the last data point due to this leverage.

As pointed out, a limitation of the two-stage hormetic model developed here is the symmetry of a hormetic zone. This assumption may lead to a positive bias as shown in the third

simulation scenario and the two applied examples. This behavior has implication with respect to public safety since overestimation of the BMD is generally more detrimental than the underestimation of the same magnitude in cancer risk assessments. To break the symmetry, we could add an additional parameter in the linear predictor as $\sum_{m=0}^{2} \beta_m \left(d_i/d_I\right)^{\beta_3}$, where x_I is the maximum experimental dose, if $I > 4$. In addition to allowing an asymmetric hormetic zone, a potential leverage effect can be reduced if $\beta_3 < 1$. This parameterization in the multistage model still allows us to partition the parameter space by the sign of β_1. The three-stage model with the linear predictor $\sum_{m=0}^{3} \beta_m d_i$ does not have this property (i.e. a single parameter determines a class of assumption), thus classification of dose-response patterns is not simple. In the applied examples, however, using the four-parameter (asymmetric) hormetic model might not be appropriate since $I = 4$ for both datasets. However, such a parameterization may be useful when data are available at more experimental doses.

If the true dose-response relationship is strictly monotonic, the parameter of interest from a regulatory perspective may still be the BMD for a given BMR. On the other hand, if the dose-response relationship is hormetic, the most beneficial dose δ satisfying $\pi_\delta = \min_{x>0} \pi_x$ is an alternative parameter of interest. In either case it is important to develop robust and efficient statistical procedures that allow for monotonicity or hormesis, and that provide information on the posterior probabilities for these two classes of models. In addition, optimal study designs for maximizing information regarding potential hormetic effects are needed. These remain focuses of our future research.

Chapter 5

Hypothesis Testing for Hormesis

In toxicology studies hormesis refers to a dose-response relationship with a stimulatory response at low doses and an inhibitory response at high doses. In Chapter 5, we focus on statistical methods for hypothesis testing for hormesis. We consider both parametric and non-parametric approaches and illustrate that parametric assumptions may lead to decreased statistical power in some cases.

5.1 Introduction

Hormesis has been extensively discussed in past literature. Calabrese and Baldwin applied a priori criteria to a large number of toxicology studies to determine whether each study provided sufficient evidence for hormesis [11, 10]. Later, additional studies evaluated hormetic effects in carcinogen risk assessments [14, 15, 19]. Although these studies highlighted the potential importance of hormesis, other authors have expressed skepticism, pointing out a lack of formality in the hypothesis testing procedures, unknown specificity and sensitivity, and potential adverse consequences of incorporating hormetic models into policy decisions

[21, 61, 42].

When attempting to distinguish hormesis from monotonicitic dose-response relationship, the importance of flexible dose-response modeling has been been discussed, and many nonmonotonic dose-response models have been proposed [58, 53, 34, 33, 32, 8, 7, 37]. Among many existing models, the parametric model proposed by Hunt and Bowman [34] classifies dose-response relationships into three classes: strictly increasing, threshold (i.e. flat below some threshold), and hormetic. This model was extended by Hunt and Rai to account for potential random effects due to litter affiliation when testing for hormetic and threshold effects in animal studies [33]. The sample size they considered was relatively large, therefore the method could be applied by relying on asymptotic results in a frequentist framework. Recently, Kim et al [37] considered a multistage model which allows monotonicity and hormesis for the estimation of a benchmark dose defined by the Environmental Protection Agency [26]. The multistage model is a member of a class of well-known models included in the widely used EPA Benchmark Dose Software.

The common property of the aforementioned parametric models is that they focus on symmetric hormetic zones only. Our proposal in this chapter is motivated by potential caveats due to model misspecification when the true hormetic zone is not symmetric. To address the issue, we propose two alternative methods. The first is based on the multistage model with a nonlinear predictor. This parametric approach is able to fit an asymmetric hormetic zone with enhanced flexibility. The second is based on a weighted average of multiple nonparametric models using Bayesian model averaging (BMA) as proposed by [50]. This nonparametric approach quantifies the posterior probability of hormesis based on the BMA framework. In this chapter we mainly focus on hypothesis testing for strict monotonicity (the null hypothesis) versus hormesis (the alternative hypothesis).

The remainder of the chapter is organized as follows. In Section 5.2 we review the aforementioned Hunt-Bowman model and propose the flexible multistage model and the non-

parametric BMA approach. In Section Section 5.3 presents the results of a simulation study to assess the performance of each approach for distinguishing between hormetic from monotonic dose-response relationships in relatively sparse data settings that are commonly encountered in toxicology studies. In Section 5.4 we apply the proposed methods to data assessing the carcinogenic effect of cadmium compounds in male rats. Section 5.5 summarizes the methods considered and discusses avenues for further research in the field.

5.2 Methods

In this section we discuss three dose-response models that allow for both monotonicity and hormesis. The first model is the Hunt-Bowman model originally proposed by Hunt and Bowman [34] in a frequentist framework, though we implement the model in a Bayesian framework to allow for potentially better performance in sparse data situations. The second model is a multistage model with a nonlinear predictor, which we refer to as the multistage model for the remainder of the chapter. The third model is a BMA-based nonparametric approach that does not require any functional form for a dose-response relationship. Under each model, we consider hypothesis testing for strict monotonicity versus hormesis.

5.2.1 Hunt-Bowman Model

Parameterization.

Let $d \geq 0$ denote the experimental dose of a potentially toxic compound. The Hunt-Bowman model describes a dose-response relationship by

$$\pi_d(\vec{\theta}) = \alpha_0 + \alpha_1 d + \alpha_2 d^2 \tag{5.1}$$

for $0 \leq d \leq \tau$ and

$$\pi_d(\vec{\theta}) = \frac{1}{1 + e^{-\beta_0 - \beta_1(d-\tau)}} \tag{5.2}$$

for $d > \tau$, where $\vec{\theta} \equiv (\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \tau)$ is a vector of unknown model parameters and $\pi_d(\vec{\theta})$ is the probability of a positive response at dose d. To force continuity in d, there are two restrictions $\alpha_2 = -\alpha_1/\tau$ and $\alpha_0 = (1+e^{\beta_0})^{-1}$. These restrictions do not guarantee continuity in the slope of a dose-response curve. For $0 \leq d \leq \tau$, Equation (5.1) can be written as

$$\pi_d(\vec{\theta}) = \left(\frac{1}{1 + e^{-\beta_0}}\right) + \alpha d - \left(\frac{\alpha}{\tau}\right) d^2, \tag{5.3}$$

where $\alpha \equiv \alpha_1$ for short, and the vector of freely varying parameters is $\vec{\theta} = (\alpha, \beta_0, \beta_1, \tau)$. This parameterization allows for a hormetic effect for $d \in (0, \tau)$ using the quadratic form in Equation (5.3) and resumes the monotonic curve by the two-parameter logistic model in Equation (5.2). In the absence of hormesis, it is equivalent to the usual two-parameter logistic model with a background risk at $d = 0$.

If $\tau = 0$, the parameter space is

$$\Theta_M = \{(\alpha, \beta_0, \beta_1, \tau) : -\infty < \beta_0 < \infty, \beta_1 > 0, \tau = 0\} \ ,$$

which does not depend on α. If $\tau > 0$ and $\alpha \neq 0$ (i.e. hormesis), the parameter space is a subset of

$$\Theta_0 = \{(\alpha, \beta_0, \beta_1, \tau) : \alpha < 0, -\infty < \beta_0 < \infty, \beta_1 > 0, \tau > 0\} \ ,$$

where the subset guarantees $\pi_d(\vec{\theta}) > 0$ for all $d \geq 0$. The sufficient condition for $\pi_d(\vec{\theta}) > 0$ for all $d \geq 0$ is the single inequality $\pi_{0.5\tau}(\vec{\theta}) > 0$ in Equation (5.3). Therefore, the parameter

space for hormesis is

$$\Theta_H = \left\{ \vec{\theta} \in \Theta_0 : \pi_{0.5\tau}(\vec{\theta}) > 0 \right\} .$$

In the presence of hormesis, τ represents the length of hormetic zone (defined as the range of doses where the probability of a toxic event is less than that observed at no dose), the magnitude of $\alpha < 0$ determines the strength of the hormetic effect, β_0 determines the background risk $\pi_0(\vec{\theta})$, and $\beta_1 > 0$ determines the rate of increasing toxicity with respect to $d \in (\tau, \infty)$. We wish to test H_0: $\tau = 0$ versus H_1: $\tau > 0$, and note that there are four freely varying parameters in Θ_H and two freely varying parameters in Θ_M. If $\tau > 0$ and $\alpha = 0$, the model generates a threshold dose-response relationship, a class not focused on in this chapter.

Prior.

To implement the Hunt-Bowman model in a Bayesian framework, it is necessary to specify a prior distribution for the model parameters. As the single value of τ (i.e. $\tau = 0$) determines the absence of hormesis, we consider a spike and slab prior for τ [35]. We let $p_0 = P(\tau > 0)$, the prior probability of hormesis. Then, the prior density function of τ is

$$f(\tau) = p_0 \, f^+(\tau) + (1 - p_0) \, \mathrm{I}_{\tau=0} \, , \tag{5.4}$$

where $\mathrm{I}_{\tau=0} = 1$ if $\tau = 0$ (zero otherwise) and $f^+(\tau)$ is a density function for $\tau > 0$.

For an interpretable prior specification on (β_0, β_1), we consider a conditional mean prior (Bedrick *et al.*, 1996). By choosing two arbitrarily doses, say d_{-2} and d_{-1}, we independently model

$$\pi_{d_i}(\beta_0, \beta_1) = \frac{e^{\beta_0 + \beta_1 d_i}}{1 + e^{\beta_0 + \beta_1 d_i}} \sim \mathrm{Beta}(r_i, s_i)$$

using the logistic link. Using Jacobian transformation, it can be shown that

$$f(\beta_0, \beta_1) = \prod_{i=-2}^{-1} \frac{\Gamma(r_i + s_i)}{\Gamma(r_i)\,\Gamma(s_i)} \left(\frac{e^{\beta_0+\beta_1 d_i}}{1+e^{\beta_0+\beta_1 d_i}}\right)^{r_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 d_i}}\right)^{s_i} . \tag{5.5}$$

By conjugacy, $r_i = y_i$ and $s_i = n_i - y_i$ are interpreted as pseudo observations at the fixed dose d_i for $i = -2, -1$. If it is necessary, we may choose a set of two fixed doses separately for the monotonic and hormetic case with separate Beta distributions.

If $\tau = 0$, we do not need to consider a prior distribution for α. Given $\tau > 0$ and β_0, a hormetic effect can be quantified as $\pi_0(\vec{\theta}) - \pi_{0.5\tau}(\vec{\theta}) = -\alpha\tau/4$. To express large uncertainty for $-\alpha\tau/4$, we may consider the uniform distribution

$$-\frac{\alpha\tau}{4} \,\Big|\, \tau, \beta_0 \sim \text{Uniform}\left(0, \frac{1}{1+e^{-\beta_0}}\right) .$$

That is,

$$f(\alpha \mid \tau, \beta_0) = \frac{\tau(1+e^{-\beta_0})}{4} , \quad \alpha \in \left(-\frac{4}{\tau(1+e^{-\beta_0})}, 0\right) . \tag{5.6}$$

From Equations (5.4), (5.5), and (5.6), the joint prior density function of $\vec{\theta} = (\alpha, \beta_0, \beta_1, \tau)$ is

$$f(\vec{\theta}) = f(\alpha \mid \tau, \beta_0)\, f(\beta_0, \beta_1)\, f(\tau) \tag{5.7}$$

for $\vec{\theta}$ in the parameter space $\Theta_M \cup \Theta_H$.

Posterior.

Let d_i denote experimental doses for $i = 0, 1, \ldots, I$ with $d_0 = 0$ being the control group and $0 < d_1 < \cdots < d_I$. Without loss of generality, we assume $d_I = 1$. Given n_i animals allocated at dose d_i, let $Y_i \sim \text{Binomial}(n_i, \pi_i)$ denote the number of toxic events with $\pi_i = \pi_{d_i}(\vec{\theta})$, the probability of a toxic event as a function of d_i given by the Hunt-Bowman model in Section 2.1.1. The likelihood function is

$$f(\vec{y} \mid \vec{\theta}) = \prod_{i=0}^{I} \frac{n_i!}{y!\,(n_i - y_i)!} \left\{\pi_{d_i}(\vec{\theta})\right\}^{y_i} \left\{1 - \pi_{d_i}(\vec{\theta})\right\}^{n_i - y_i} \tag{5.8}$$

for given data $\vec{y} = (y_0, y_1, \ldots, y_n)$.

By assuming $\tau \sim \text{Uniform}(d_0, d_I)$ in the presence of hormesis, we let $f^+(\tau) = 1$ for $\tau \in (0, 1)$ in Equation (5.4). From Equations (5.7) and (5.8), the joint posterior density function is

$$\begin{aligned} f(\vec{y} \mid \vec{\theta}) &= \frac{f(\vec{y} \mid \vec{\theta})\, f(\vec{\theta})}{\int_{\Theta_M \cup \Theta_H} f(\vec{y} \mid \vec{\theta})\, f(\vec{\theta})\, d\vec{\theta}} \\ &\propto \left\{ \frac{p_0 \tau (1 + e^{-\beta_0})}{4}\, \mathrm{I}_{\tau > 0} + (1 - p_0)\, \mathrm{I}_{\tau = 0} \right\} \prod_{i=-2}^{I} c_i \left\{\pi_{d_i}(\vec{\theta})\right\}^{y_i} \left\{1 - \pi_{d_i}(\vec{\theta})\right\}^{n_i - y_i} \end{aligned}$$

for $\vec{\theta}$ respecting all constraints in the parameter space $\Theta_M \cup \Theta_H$, where c_i is the appropriate constant for $i = -2, \ldots, I$ and $y_i = r_i$ and $n_i = s_i + r_i$ for $i = -2, -1$. Particularly, in addition to the boundary set in $\Theta_M \cup \Theta_H$, we condition on $\tau \in [0, 1]$ and force α to respect the bounds given in Equation (5.6). Then, posterior inference for the hypothesis of hormesis can be based on $P(H_0 \mid \vec{y}) = P(\tau = 0 \mid \vec{y})$ and $P(H_1 \mid \vec{y}) = P(\tau > 0 \mid \vec{y})$.

5.2.2 Multistage Model

Parameterization.

Let $d \geq 0$ be a dose as in the previous section. The linearized multistage model describes a monotonic dose-response relationship by

$$\pi_d(\vec{\gamma}) = 1 - e^{-\sum_{m=0}^{M} \gamma_m d^m}$$

with nonnegative coefficients [2, 1, 20]. We let $M = 2$ with an additional parameter γ_3 such that

$$\pi_d(\vec{\theta}) = 1 - e^{-\sum_{m=0}^{2} \gamma_m d^{m\gamma_3}} , \tag{5.9}$$

and we allow $\gamma_1 < 0$ to model a hormetic effect at low dose. Bogen and Kim et al considered the linear predictor (i.e. $\gamma_3 = 1$) for hormesis which allows a symmetric hormetic zone [8, 37]. By the nonlinear predictor in Equation (6.2), we are able to model an asymmetric hormetic zone with a smooth transition from a hormetic zone to a monotonic zone as opposed to the Hunt-Bowman model.

We can partition the four-dimensional parameter space for $\vec{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ into two disjoint spaces. The monotonic parameter space is given by

$$\Gamma_M = \{\vec{\gamma} : \gamma_0 > 0, \gamma_1 \geq 0, \gamma_2 > 0, \gamma_3 > 0\} ,$$

and the hormetic parameter space is given by

$$\Gamma_H = \left\{\vec{\gamma} : \gamma_0 > 0, \gamma_1 < 0, \gamma_2 > 0, \gamma_3 > 0, \gamma_1^2 < 4\gamma_0\gamma_2\right\} .$$

The last inequality guarantees $\pi_{d^*} > 0$, where

$$d^* = d^*(\vec{\gamma}) = \left(-\frac{\gamma_1}{2\gamma_2}\right)^{1/\gamma_3} \tag{5.10}$$

is the nadir of the dose-response curve (i.e. the most beneficial dose). The sufficient condition for $\pi_d(\vec{\gamma}) > 0$ for all $d \geq 0$ is $\sum_{m=0}^{2} \gamma_m (d^*)^{m\gamma_3} > 0$. Under this model, the length of the hormetic zone is

$$d^{**} = d^{**}(\vec{\gamma}) = \left(-\frac{\gamma_1}{\gamma_2}\right)^{1/\gamma_3}, \tag{5.11}$$

where $\pi_0(\vec{\gamma}) = \pi_{d^{**}}(\vec{\gamma})$ and $\pi_d(\vec{\gamma}) < \pi_0(\vec{\gamma})$ for all $d \in (0, d^{**})$.

One distinction from the Hunt-Bowman model is that all four parameters are present in both monotonic case and hormetic case. However, the nonlinear predictor may cause over-fitting, particularly when the number of experimental doses is small. To avoid fitting a hormetic curve to monotonic empirical points, we may impose some restrictions on Γ_H. In particular, we may be reluctant to allow $d^* < d_1$ and $d^{**} < d_2$ based on an experimental design with a small number of experimental doses. In addition, we may not be practically interested in extremely small hormetic effects (i.e. minimal hormetic effect). For example, if we measure a hormetic effect by the odds ratio comparing the odds of toxicity at doses $d = 0$ and $d = d^*$,

$$\eta(\vec{\gamma}) = \frac{\pi_0(\vec{\gamma})}{1 - \pi_0(\vec{\gamma})} \frac{1 - \pi_{d^*}(\vec{\gamma})}{\pi_{d^*}(\vec{\gamma})}, \tag{5.12}$$

we may be interested in a hormetic case where $\eta(\vec{\gamma})$ exceeds some threshold. To this end, we define a restricted hormetic zone as

$$\Gamma_H^{(R)} = \{\vec{\gamma} \in \Gamma_H : d^*(\vec{\gamma}) \geq d_L^*, d^{**}(\vec{\gamma}) \geq d_L^{**}, \eta(\vec{\gamma}) \geq \eta_L\}$$

for some $d_L^* \geq d_1$, $d_L^{**} \geq d_2$, and $\eta_L > 1$. In other words, $\Gamma_H^{(R)}$ is the hormetic zone of practical importance, and $\Gamma_H - \Gamma_H^{(R)}$ becomes the hormetic zone of indifference.

The role of each parameter in the multistage model is as follows: The background risk $\pi_0(\gamma_0) = 1 - e^{-\gamma_0}$ is completely determined by γ_0, and the sign of γ_1 determines the class of a dose-response relationship. If $\vec{\gamma} \in \Gamma_M$, we have $\gamma_1 \geq 0$. In this case, γ_2 and γ_3 provide additional flexibility in a strictly monotonic dose-response curve. If $\vec{\gamma} \in \Gamma_H$, we have $\gamma_1 < 0$. In this case, γ_3 controls the shape of the hormetic zone. Defining $\rho \equiv d^*/d^{**} \in (0, 1)$, the ratio of the most beneficial dose to the length of the hormetic zone, we have $\gamma_3 = \log(0.5) \,/\, \log(\rho)$ from Equations (5.10) and (5.11). Then, $\gamma_3 = 1$ (i.e. $\rho = 0.5$) if the hormetic zone is symmetric and $\gamma_3 \neq 1$ (i.e. $\rho \neq 0.5$) if the hormetic zone is asymmetric.

A test of the hypothesis $H_0 : \vec{\gamma} \in \Gamma_M$ (i.e. strict monotonicity) versus $H_1 : \vec{\gamma} \in \Gamma_H$ (i.e. hormesis) is equivalent to testing H_0: $\gamma_1 \geq 0$ versus H_1: $\gamma_1 < 0$ for $\vec{\gamma} \in \Gamma_M \cup \Gamma_H$. If we consider the restricted hormetic zone $\Gamma_H^{(R)}$, the sign of γ_1 distinguishes between strict monotonicity and hormesis in the same manner.

The parameterization in Equation (6.2) dose not allow a threshold case, but an arbitrary value of $\gamma_3 > 1$ generates a threshold-like curve. If one wants to classify a dose-response relationship into three classes including a threshold case, we may partition $\Gamma_M \cup \Gamma_H$ into three regions, the zone of strict monotonicity, the zone of hormesis, and the zone of indifference. However, the choice of boundaries for defining the corresponding parameter space regions is not trivial, and it is also necessary to consider scientific and practical importance of distinguishing the threshold case from the monotonic and hermetic cases.

Prior.

Unlike the prior specification under the Hunt-Bowman model, we do not need a spike and slab prior under the multistage model. Instead, we specify a conditional joint prior distribution

for $(\gamma_0, \gamma_1, \gamma_2)$ given γ_3 using a conditional mean prior. Operationally, we can arbitrarily select three doses, say $d_{-3} < d_{-2} < d_{-1}$, and specify three independent Beta priors as

$$\pi_{d_i}(\vec{\gamma}) = 1 - e^{-\sum_{m=0}^{2} \gamma_m d_i^{m\gamma_3}} \sim \text{Beta}(t_i, u_i)$$

for $i = -3, -2, -1$. To transform from the three independent Beta distributions to the conditional joint distribution of $(\gamma_0, \gamma_1, \gamma_2)$ given γ_3, it can be shown that the determinant of the Jacobian matrix is

$$\det(J) = (c_{-1,0} - c_{-2,0} + c_{-2,-1}) \prod_{i=-3}^{-1} \{1 - \pi_{d_i}(\vec{\gamma})\} \ ,$$

where $c_{j,k} = (d_j d_k)^{\gamma_3}(d_k - d_j)$. By letting $c(\gamma_3) = | \, c_{-1,0} - c_{-2,0} + c_{-2,-1} \, |$ the conditional joint prior density function of $(\gamma_0, \gamma_1, \gamma_2)$ given γ_3 is

$$f(\gamma_0, \gamma_1, \gamma_2 \mid \gamma_3) = c(\gamma_3) \prod_{i=-3}^{-1} \frac{\Gamma(t_i + u_i)}{\Gamma(t_i)\,\Gamma(u_i)} \left(1 - e^{-\sum_{m=0}^{2} \gamma_m d_i^{m\gamma_3}}\right)^{t_i - 1} \left(e^{-\sum_{m=0}^{2} \gamma_m d_i^{m\gamma_3}}\right)^{u_i} . \tag{5.13}$$

Based on a prior belief about γ_3, which determines the shape of a hormetic zone as described in the previous section, we let $\gamma_3 \sim f(\gamma_3)$ for $\gamma_3 > 0$. Then the joint prior density function is given by $f(\vec{\gamma}) = f(\gamma_3)\, f(\gamma_0, \gamma_1, \gamma_2 \mid \gamma_3)$. Similar to the prior specification described in Section 5.2.1, we consider $t_i - 1 = y_i$ and $u_i = n_i - y_i$ as pseudo observations at the fixed dose d_i for $i = -3, -2, -1$.

Posterior.

Using the same form of the likelihood function in Equation (5.8), the joint posterior density function of $\vec{\gamma}$ is given by

$$f(\vec{\gamma} \mid \vec{y}) \propto f(\gamma_3)\, c(\gamma_3) \prod_{i=-3}^{I} \left(1 - e^{-\sum_{m=0}^{2} \gamma_m d_i^{m\gamma_3}}\right)^{y_i} \left(e^{-\sum_{m=0}^{2} \gamma_m d_i^{m\gamma_3}}\right)^{n_i - y_i}$$

for $\vec{\gamma} \in \Gamma_M \cup \Gamma_H$ for the unrestricted hormetic zone or $\vec{\gamma} \in \Gamma_M \cup \Gamma_H^{(R)}$ for the restricted hormetic zone. As such, posterior inference for the hypothesis of hormesis may be based on $P(\gamma_1 < 0 \mid \vec{y}) = 1 - P(\gamma_1 \geq 0 \mid \vec{y})$.

5.2.3 Nonparametric Models with Bayesian Model Averaging

Parameterization.

Suppose we have $I + 1$ experimental doses $d_0 < d_1 < \cdots < d_I$, where $d_0 = 0$ is the control group. We let M_j denote the j^{th} dose-response model and $\pi_{ij} = \pi_{d_i,j}$ denote the probability of a toxic event at dose d_i under M_j for $i = 0, 1, \ldots, I$. We do not introduce any mathematical relationship between d_i and π_{ij}. Instead we consider I nonparametric models, say $M_0, \ldots, M_{I-1}$. Without loss of generosity, let M_j denote a model such that $\min(\pi_{0j}, \pi_{1j}, \ldots, \pi_{Ij}) = \pi_{jj}$ for $j = 0, 1, \ldots, I - 1$. By restricting that a dose-response curve cannot have a negative slope once it passes its nadir, $M_1, \ldots, M_{I-1}$ represent hormetic models with the nadirs at $d_1, \ldots, d_{I-1}$, respectively, and M_0 represents a strictly monotonic model. It is difficult to believe that a nadir of a hormetic zone is exactly located at some experimental dose d_i, but this formulation is based on convenience. In this nonparametric modeling, the hypothesis testing for monotonicity versus hormesis is equivalent to H_0: M_0 versus H_1: $\bigcup_{j=1}^{I-1} M_j$.

Prior.

We let $\vec{\pi}_j = (\pi_{0j}, \pi_{1j}, \ldots, \pi_{Ij})$ denote the vector of parameters under M_j for $j = 0, 1, \ldots, I-1$. For a joint prior distribution of $\vec{\pi}_j$, we consider a series of conditional truncated Beta distributions. We say $\pi \sim \text{TB}(a, b, l, r)$ if the probability density function is

$$f(\pi) = \frac{g(\pi)}{\int_l^r g(\pi)\, d\pi},$$

where

$$g(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \pi^{a-1}\,(1-\pi)^{b-1}$$

for $\pi \in (l, r)$ and $g(\pi) = 0$ for $\pi \notin (l, r)$. For the control dose group, we assume a Beta distribution which is equivalent to $\pi_{0j} \sim \text{TB}(a_{0j}, b_{0j}, 0, 1)$, and this prior distribution may be constant over M_j based on the prior knowledge regarding the prevalence of a toxic event in the absence of the toxin. If $\pi_{i+1,j}$ shall be greater than π_{ij}, we assume

$$\pi_{i+1,j} \mid \pi_{ij} \sim \text{TB}(a_{ij}, b_{ij}, \pi_{ij}, r_{ij})$$

for some $r_{ij} \in (\pi_{ij}, 1)$. If $\pi_{i+1,j}$ shall be smaller than π_{ij}, we assume

$$\pi_{i+1,j} \mid \pi_{ij} \sim \text{TB}(a_{ij}, b_{ij}, l_{ij}, \pi_{ij})$$

for some $l_{ij} \in (0, \pi_{ij})$. To accept a wide range of dose-response curves, we may let $l_{ij} = 0$ and $r_{ij} = 1$. By the multiplication rule, the joint prior density function of $\vec{\pi}_j$ under M_j is

$$f(\vec{\pi}_j \mid M_j) = f(\pi_{0j}) \prod_{i=1}^{I} f(\pi_{ij} \mid \pi_{i-1,j})$$

for $j = 0, 1, \ldots, I-1$.

In addition to the prior specification on each $\vec{\pi}_j$, we assign the prior model probability $P(M_j) > 0$ such that $\sum_{j=0}^{I-1} P(M_j) = 1$. This probability assignment shoud reflect the prior plausibility of each hypothesis as $P(H_0) = P(M_0)$ and $P(H_1) = \sum_{j=1}^{I-1} P(M_j)$. For example, $P(M_0) = 1/2$ and $P(M_j) = \frac{1}{I-1}$ for $j = 1, \ldots, I-1$ reflect the same degree of prior belief for monotonicity and hormesis, and they reflect the same degree of prior belief for each nadir conditioning on hormesis.

Posterior.

The posterior probabilities of interest are

$$P(H_0 \mid \vec{y}) = P(M_0 \mid \vec{y})$$

and

$$P(H_1 \mid \vec{y}) = \sum_{j=1}^{I-1} P(M_j \mid \vec{y})\,.$$

Appealing to Bayesian model averaging (BMA), each posterior model probability is

$$P(M_j \mid \vec{y}) = \frac{f(\vec{y} \mid M_j)\, P(M_j)}{\sum_{j=0}^{I-1} f(\vec{y} \mid M_j)\, P(M_j)}\,,$$

where $f(\vec{y} \mid M_j)$ is the marginal likelihood function under M_j. That is,

$$f(\vec{y} \mid M_j) = \int_{\Omega_j} f(\vec{y} \mid \vec{\pi}_j, M_j)\, f(\vec{\pi}_j \mid M_j)\, d\vec{\pi}_j$$

where Ω_j is a subset of $(0,1) \times \cdots \times (0,1)$ based on the truncations under M_j and

$$f(\vec{y} \mid \vec{\pi}_j, M_j) = \prod_{i=0}^{I} \frac{n_i!}{y_i!\,(n_i - y_i)!}\ (\pi_{ij})^{y_i}\ (1 - \pi_{ij})^{n_i - y_i}$$

is the likelihood function under M_j.

5.3 Simulations

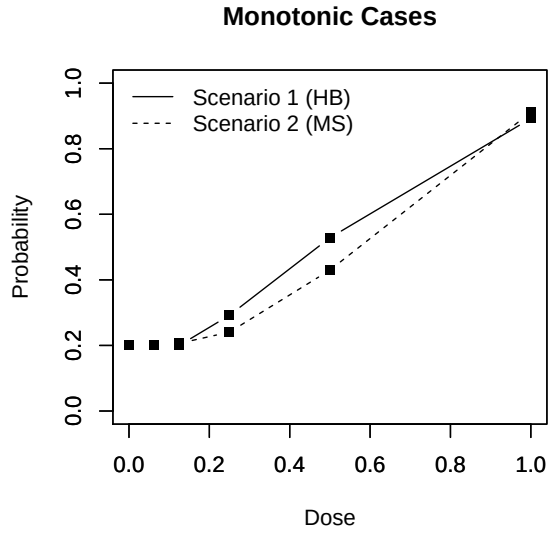
In this section, we investigate sensitivity and specificity of the Hunt-Bowman model, the multistage model, and the BMA-based nonparametric method in the context of hypothesis testing for hormesis.

5.3.1 Scenarios

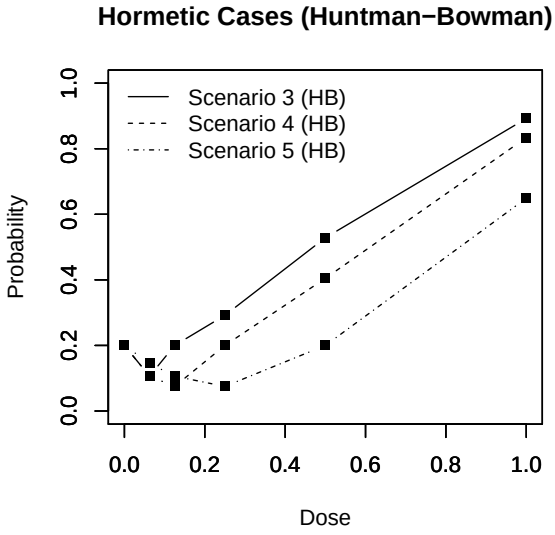
We consider two monotonic scenarios and seven hormetic scenarios with various shapes of a hormetic zone. Scenario 1 is a threshold case which is generated under the Hunt-Bowman model with $\vec{\theta} = (\alpha, \beta_0, \beta_1, \tau) = (0, -1.386, 4, 0.125)$. Scenario 2 is a strictly monotonic case which is generated under the multistage model with $\vec{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (0.223, 0, 1.5, 2)$. The two monotonic scenarios are presented in Figure 5.1a. Scenarios 1 and 2 serve as references when we evaluate specificity (i.e. concluding H_0 when H_0 is true) and sensitivity (i.e. concluding H_1 when H_1 is true) in the context of hypothesis testing. A model tending to yield a large $P(H_1 \mid \vec{y})$ under a monotonic scenario tends to yield a large $P(H_1 \mid \vec{y})$ under a hormetic scenario as well, therefore we need some monotonic scenarios for comparison.

Scenarios 3 to 5 are hormetic cases generated under the Hunt-Bowman model. The respective true parameters are $\vec{\theta} = (-3, -1.386, 4, 0.125)$, $\vec{\theta} = (-2, -1.386, 4, 0.25)$, and $\vec{\theta} = (-1, -1.386, 4, 0.5)$ so that the hormetic zone is shortest in Scenario 3 and longest in Scenario 5 among the three scenarios. All three hormetic zones are symmetric as shown in Figure 5.1b. Scenarios 6 and 7 are hormetic cases generated under the multistage model. The respective parameter values are $\vec{\gamma} = (0.223, -1.386, 2.773, 1)$ and $\vec{\gamma} = (0.223, -2.045, 8.180, 1)$. By fixing $\gamma_3 = 1$, the hormetic zones are symmetric, and the hormetic zone in Scenario 6

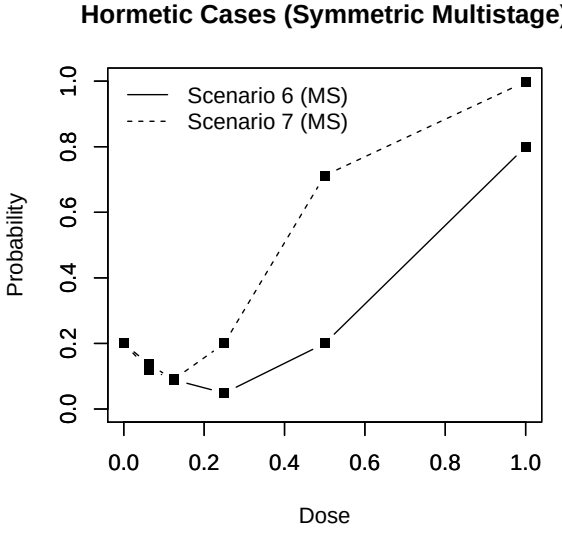
Figure 5.1: Simulation scenarios.



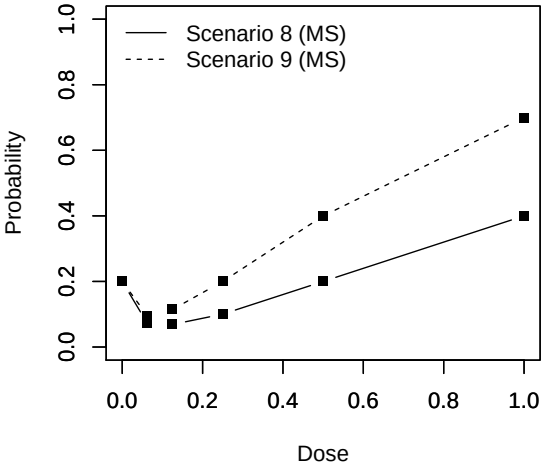
(a) Scenarios 1 and 2



(b) Scenarios 3 to 5



(c) Scenarios 6 and 7



(d) Scenarios 8 and 9

is shorter than the hormetic zone in Scenario 7 as shown in Figure 5.1c. Scenarios 8 and 9 are hormetic cases generated under the multistage model as well. The respective parameter values are $\vec{\gamma} = (0.223, -0.827, 1.145, 0.431)$ and $\vec{\gamma} = (0.223, -0.981, 1.962, 0.5)$. Since $\gamma_3 \neq 1$, the hormetic zones are asymmetric as shown in Figure 5.1d.

For each scenario, six experimental doses are geometrically spaced at $d_0 = 0$, $d_1 = 1/16$, $d_2 = 1/8$, $d_3 = 1/4$, $d_4 = 1/2$, and $d_5 = 1$, and $n_i = 50$ animals are allocated at each experimental dose which sum to the sample size of $\sum_{i=0}^{5} n_i = 300$.

5.3.2 Models Compared

We compare five different models with respect to the performance of hypothesis testing for hormesis. The first model is the Hunt-Bowman model which is denoted by HB. The second model is the multistage model with the unrestricted parameter space $\Gamma_M \cup \Gamma_H$ which is denoted by MS, and the third model is the multistage model with the restricted parameter space $\Gamma_M \cup \Gamma_H^{(R)}$ which is denoted by MSR. The MSR model rules out minimal hormetic effects by setting the lower bounds $d_L^* = d_1 = 1/16$, $\eta_L^* = d_2 = 1/8$, and $\gamma_L = 1.2$ in Equation (5.12). The fourth and fifth models are based on the nonparametric approach with BMA. We elicit vague priors in the fourth model and denote it by BMA_V, and we elicit relatively strong priors in the fifth model and denoted it by BMA_S (see Section 5.3.3 for detail explanation).

5.3.3 Priors

To elicit large uncertainty under the HB model, we specify $f^+(\tau) = 1$ for $\tau \in (0, 1)$ and $p_0 = 0.5$ in Equation (5.4). In other words, we express $P(H_0) = P(H_1)$ and the length of hormetic zone can be anywhere within the experimental range before observing data.

Then, we specify $(d_{-2}, r_{-2}, s_{-2}) = (0.2, 1.01, 1.10)$ and $(d_{-1}, r_{-1}, s_{-1}) = (0.8, 1.10, 1.01)$ in Equation (5.5). This specification adds small numbers of pseudo observations at $d_{-2} = 0.2$ and $d_{-1} = 0.8$.

Similarly, to elicit large uncertainty under the MS and MSR models, we a priori assume $(d_j, t_j, u_j) = (0.2, 1.01, 1.10), (0.5, 1.05, 1.05), (0.8, 1.10, 1.01)$ for $j = -3, -2, -1$ in Equation (5.13). This specification also adds small numbers of pseudo observations at the three chosen doses. Then we elicit a nearly flat prior on γ_3 using the $\Gamma(0.001, 0.001)$ prior.

For the BMA_V model (V stands for "vague"), we specify $(a_{ij}, b_{ij}) = (1.01, 1.10)$ for $i = 0, 1, 2, 3$, $(a_{4j}, b_{4j}) = (1.05, 1.05)$, and $(a_{5j}, b_{5j}) = (1.10, 1.01)$. This vague prior is comparable to the prior specifications in the parametric models with large uncertainty. For the BMA_S model (S stands for "strong"), we specify $(a_{ij}, b_{ij}) = (2.64, 7.55)$ for $i = 0, 1, 2, 3$, $(a_{4j}, b_{4j}) = (1.53, 1.53)$, and $(a_{5j}, b_{5j}) = (5.38, 1.49)$. This strong prior reasonably well assumes the trajectory of the dose-response curves in the simulation all scenarios. For both BMA_V and BMA_S, we have one monotonic model M_0 and four hormetic models M_i with the nadirs at d_i for $i = 1, 2, 3, 4$. For prior model probabilities we let $P(H_0) = P(H_1) = 1/2$ and $P(M_i) = 1/8$ for $i = 1, 2, 3, 4$. When $\pi_{i+1,j} > \pi_{ij}$, we let $\pi_{i+1,j} \in (\pi_{ij}, 1)$ under the truncated Beta distribution. When $\pi_{i+1,j} < \pi_{ij}$, we let $\pi_{i+1,j} \in (0, \pi_{ij})$.

5.3.4 Receiver Operating Characteristic Curve in Hypothesis Testing

Suppose we conclude H_1 when $P(H_1 \mid \vec{y}) \geq q$ for some q. By repeating simulated trials and computing $P(H_1 \mid \vec{y})$ under a null scenario and an alternative scenario, we can plot one minus specificity on the x-axis and sensitivity on the y-axis by varying the decision threshold q from zero to one. In our setting, sensitivity is the probability of concluding H_1 under an alternative scenario (Scenarios 3 to 9), and specificity is the probability of concluding

H_0 under a null scenario (Scenario 1 and 2). This plot is known as a receiver operating characteristic (ROC) curve in the context of hypothesis testing, and a large area under the curve indicates a plausible operating characteristic. Since we have two null scenarios, we can obtain two areas under ROC curves for each hormetic scenario. When Scenario 1 serves as the reference (which is generated under HB), the area is denoted by A_1. When Scenario 2 is serves as the reference (which is generated under MS), the area is denoted by A_2.

5.3.5 Results

The simulation results are summarized in Table 5.1. In the tables, $E\left[P(H_1 \mid \vec{y}_n)\right]$ denotes the average posterior probability for hormesis, and A_1 and A_2 denote the area under the ROC curve in each hormetic scenario when Scenario 1 and Scenario 2 served as the reference for a monotonic scenario, respectively. In the two monotonic scenarios, the HB model, BMA_V model and BMA_S model tended to yield relatively large $P(H_1 \mid \vec{y})$ on average when compared to the MS and MSR models. In Scenario 1, we obtained $E\left[P(H_1 \mid \vec{y}_n)\right] = 0.708$, 0.334, 0.167, 0.816, and 0.768 in the HB, MS, MSR, BMA_V and BMA_S models, respectively. In Scenario 2, the respective results were 0.860, 0.348, 0.201, 0.846, and 0.813. By the restricted hormetic zone, the MSR model yielded smaller $P(H_1 \mid \vec{y}_n)$ than the MS model on average in both monotonic scenarios.

In the three hormetic scenarios generated by the HB model, Scenarios 3 to 5, the BMA_V and BMA_S models outperformed the three parametric models based on both A_1 and A_2. Despite the true model was the HB model, the performances of MS and MSR were comparable to the HB model (see Table 5.1).

In the next two hormetic scenarios generated by the MS model with symmetric hormetic zones, Scenarios 6 and 7, the performances of BMA_S and BMA_V were robust yielding fairly large A_1 and A_2. Again, the outperformance was regardless of the amount of prior informa-

Table 5.1: Simulation results

Scenario	Truth	Length of Hormetic Zone	Model	$E\left[P(H_1 \mid \vec{y})\right]$	A_1 (relative to HB)	A_2 (relative to MS)
3	HB	1/8	HB	0.809	0.633	0.375
			MS	0.434	0.609	0.580
			MSR	0.229	0.501	0.458
			BMA_V	0.914	0.650	0.579
			BMA_S	0.883	0.656	0.572
4	HB	1/4	HB	0.943	0.860	0.653
			MS	0.668	0.850	0.842
			MSR	0.575	0.818	0.827
			BMA_V	0.994	0.926	0.887
			BMA_S	0.994	0.957	0.925
5	HB	1/2	HB	0.989	0.966	0.881
			MS	0.832	0.962	0.968
			MSR	0.792	0.922	0.959
			BMA_V	0.999	0.968	0.943
			BMA_S	1.000	0.987	0.971
6	MS	1/4	HB	0.998	0.992	0.967
			MS	0.876	0.972	0.978
			MSR	0.850	0.933	0.971
			BMA_V	1.000	0.974	0.952
			BMA_S	1.000	0.988	0.973
7	MS	1/2	HB	0.987	0.960	0.866
			MS	0.610	0.783	0.771
			MSR	0.545	0.795	0.799
			BMA_V	0.988	0.905	0.862
			BMA_S	0.989	0.934	0.892
8	MS	1/2	HB	0.973	0.913	0.715
			MS	0.877	0.975	0.981
			MSR	0.831	0.928	0.964
			BMA_V	0.999	0.970	0.946
			BMA_S	1.000	0.988	0.974
9	MS	1/4	HB	0.902	0.777	0.521
			MS	0.694	0.880	0.877
			MSR	0.520	0.781	0.784
			BMA_V	0.985	0.878	0.828
			BMA_S	0.986	0.922	0.876

tion we considered. The areas A_1 and A_2 in the HB models were comparable to the areas in the BMA_S model, and the HB model performed better than the MS and MSR models in Scenario 7. The MS or MSR model did not yield the largest A_1 and A_2 among the five models despite Scenarios 6 and 7 were generated under the MS model (see Table 5.1).

In the last two hormetic scenarios generated by the MS model with asymmetric hormetic zones, Scenarios 8 and 9, the performances of BMA_S and BMA_V were also shown to be robust with respect to both A_1 and A_2. On the other hand, the HB model showed the impact

of model misspecification in the asymmetric cases. The areas A_1 and A_2 were consistently smallest among the five models in these two scenarios. The MS model exhibited the plausible results, but it was not substantially superior to the nonparametric models despite the scenarios belonged to its own parameterization (see Table 5.1).

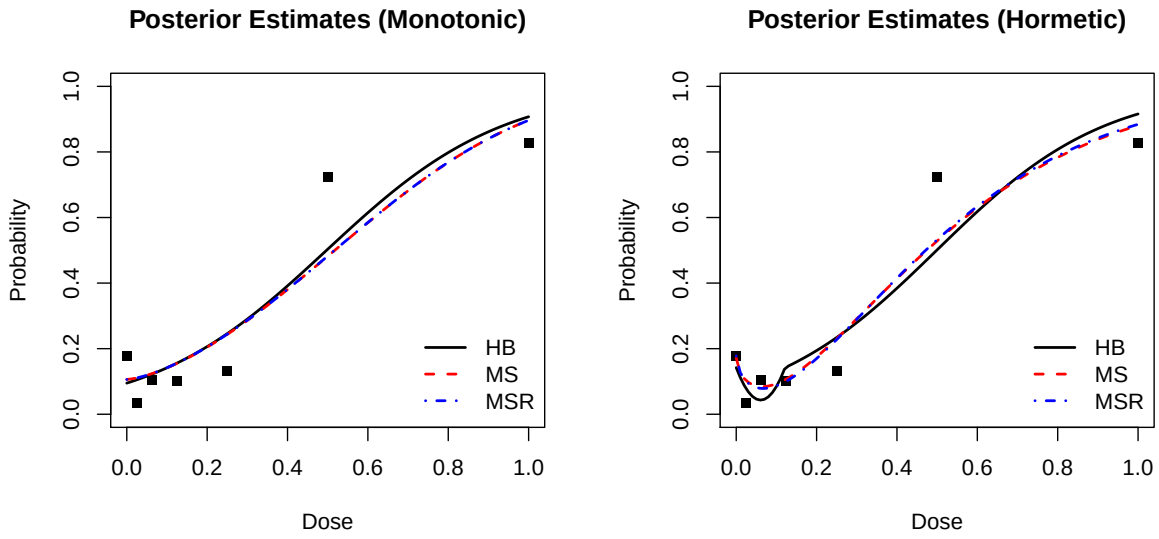
In summary, the BMA-based models which require the least assumptions about a dose-response curve showed robust results across all scenarios considered. It is difficult to compare the superiority between the HB and MS models when a true hormetic zone is symmetric. The HB model well tolerated model misspecification as long as a hormetic zone was symmetric (Scenarios 6 and 7), but it did not when a hormetic zone was asymmetric (Scenarios 8 and 9).

We repeated the same set of simulation scenarios with the smaller sample size of $n_i = 30$ for each dose group. The estimated areas A_1 and A_2 were generally smaller due to the reduced amount of data, but the relative operating characteristics were preserved.

5.4 Applications

Cadmium compounds have been known to be associated with human prostate and renal cancers. Waalkes et al studied cadmium carcinogenesis by injecting one of seven experimental doses into male rates [63]. The seven dose groups were $(0, 1, 2.5, 5, 10, 20, 40)$ in μmol/kg. By dividing each dose by the maximum dose 40 μmol/kg, the experimental doses transforms to $(0, 0.025, 0.625, 0.125, 0.25, 0.5, 1)$, respectively. This experimental doses are similar to Scenarios 8 and 9 in Section 5.3 with the additional dose $d_1 = 0.025$ between $d_0 = 0$ and $d_2 = 0.625$. We focus on the development of testicular tumors as an outcome of interest among multiple toxic outcomes measured in the study. Therefore, the respective sample sizes and observed numbers of events at the experimental doses were (45, 30, 29, 30, 30,

Figure 5.2: Estimated dose-response curves using the posterior means



(a) Monotonic

(b) Hormetic

29, 29) and (8, 1, 3, 3, 4, 21, 24), respectively. The respective observed proportions of the incidence were (0.178, 0.033, 0.103, 0.100, 0.133, 0.724, 0.828). Based on this empirical trend, the possibility of hormesis was extensively discussed in Zapponi and Marcello [71].

For the BMA model, we used similar vague priors used for the BMA_{V} model in Section 5.3.3. We specified $P(H_0) = P(H_1)$ with the uniform $P(M_j)$ for $j = 1, \ldots, 5$. For the hyper-parameters in truncated Beta distributions, we chose $(a_{ij}, b_{ij}) = (1.01, 1.10)$ for $i = 0, \ldots, 4$, $(a_{5j}, b_{5j}) = (1.05, 1.05)$, and $(a_{6j}, b_{6j}) = (1.10, 1.01)$ to express large uncertainty under all M_j. Given the data, we estimated $P(H_1 \mid \vec{y}) \approx 1$. When we altered the hyper-parameters to $(a_{ij}, b_{ij}) = (2.64, 7.55)$ for $i = 0, \ldots, 4$, $(a_{5j}, b_{5j}) = (1.53, 1.53)$, and $(a_{6j}, b_{6j}) = (5.38, 1.49)$, we obtained $P(H_1 \mid \vec{y}) = 0.998$. We attempted various strong priors, and they altered the posterior probabilities in various degrees while they maintained $P(H_1 \mid \vec{y})$ closed to one.

For the three parametric models, HB, MS, and MSR, we used the same priors as described in Section 5.3.3. The fitted dose-response curves are presented in Figure 5.2. Figure 5.2a presents the fitted curves conditioning on monotonicity, and Figure 5.2b presents the fitted curves conditioning on hormesis. The posterior probability for hormesis was 0.877, 0.821, and 0.851 under HB, MS, and MSR, respectively. We did not observe sudden changes in $P(H_1 \mid \vec{y})$ from various priors we attempted in a reasonable range.

For additional analyses using the parametric models, we estimated the length of hormetic zone which is denoted by τ under the HB model and d^{**} under the MS and MSR models as defined in Equation (5.11). The posterior mean was 0.122 with 95% credible interval (CI) of (0.030, 0.272) under HB, 0.183 with 95% CI of (0.042, 0.315) under MS, and 0.200 with 95% CI of (0.103, 0.319). The estimated hormetic zone was the shortest under the HB model potentially due to model misspecification. The empirically shown hormetic zone in Figure 5.2b does not seem to be symmetric. The difference in the estimation under MS and MSR was small, but the estimated hormetic zone was slightly longer under MSR by disregarding too short hormetic zone and minimal hormetic effect in $\Gamma_H^{(R)}$. In addition, the posterior interval was shorter under MSR than under MS. When we compare to the empirical points where the observed proportion at $d_4 = 0.25$ (0.133) is still smaller than the observed proportion at the control dose $d_0 = 0$ (0.178), the three model-specific posterior means seem to underestimate the length of hormetic zone. In particular, the 95% CI from the HB model barely covers $d_4 = 0.25$.

5.5 Discussions

We compared the three different models for hypothesis testing for hormesis. The Hunt-Bowman model is inflexible in a sense it cannot model an asymmetric hormetic zone, but it is flexible in a sense that it can fit a hormetic zone and a monotonic zone separately. The

multistage model is flexible in a sense that it can approximate various shapes of hormetic zones, but it is inflexible in a sense that it needs to model the two zones simultaneously. The nonparametric approach using BMA is plausible in both aspects.

When we compare the simulation results with 50 animals per dose level and the simulation results with 30 animals per dose level (six dose groups total), the BMA-based nonparametric approach seems most appealing among the models we compared here. If the BMA method is our choice, however, we need to carefully choose a joint prior even in a lack of prior information because marginal likelihood is sensitive to a prior specification. When we have sufficiently large sample sizes, we may gain efficiency by imposing further restrictions on unreasonable prior dose-response paths based on available prior knowledge.

Though we could not exhaust all possible dose-response curves in the simulation studies, we could learn that the parametric approaches may not be reliable particularly under poorly designed studies. A parametric approach is heavily influenced by leverage points, and experimental designs with too much information at high doses may be an issue for detecting hormesis under a parametric model. We may be able to create a scenario with an assymetric hormetic zone such that the MS model performs poorer than the HB model. Dette et al proposed optimal experimental designs for hormesis studies, but an optimal design is not necessarily optimal under model misspecification [23]. To reduce the impact of model misspecification in parametric approach and to gain more efficiency in nonparametric approach, an efficient experimental design in a nonparametric framework is a focus of our future research.

Chapter 6

Experimental Designs for Detecting Hormesis

Though scientists have been interested in hormesis, most existing studies from poor experimental design with regards to testing for hormesis. We now turn our focus to situations where we are able to control the distribution of the number of observations at each experimental dose. In Chapter 6, we develop and apply Bayesian optimal designs and also consider a new robust design based on Bayesian decision theory.

6.1 Introduction

Most existing datasets have not been collected in order to efficiently test for the presence of hormesis. In particular, many studies suffer from a lack of experimental doses that are equally spaced over the dose range, and allocate observations uniformly over these points. As with any other inferential procedure, a large sample size and a large number of experimental doses are preferable for maximising one's ability to detect hormesis. However, in most practical

cases logistics demand that risk assessment studies utilise a relatively small fixed sample size and a fixed number of experimental doses. Given this constraint, statistical design strategies to improve sensitivity for detecting hormesis are needed.

Calabrese and Baldwin [11] established quantitative criteria to categorize datasets from no evidence to high evidence for hormesis. Their method was based on an estimated hormetic zone, the number of experimental doses below an estimated hormetic zone, statistical significance of stimulatory responses, its practical significance and replicability of the findings. Extensive discussions regarding hormetic effects of toxic agents have followed [10, 14, 15, 19]. On the other hand, some investigators have pointed out a lack of formal procedures (i.e. statistical hypothesis testing) and possible caveats for claiming hormesis as the a priori dose-response relationship in risk assessments [21, 61, 42].

Multiple dose-response models based on various statistical frameworks have been developed to model a non-monotonic dose-response relationship [32, 33, 53, 8, 7, 37]. While many statistical methods have mainly focused on analyzing observed data, Dette et al. [23] proposed Bayesian optimal experimental designs for testing for hormesis. In their approach, a decision (i.e. allocation of doses for assessing response) is made prior to the initiation of the experiment rather than sequentially updating the dose allocation criteria periodically as observed responses are obtained. The method proposed by Dette et al. [23] considered a criterion-robust design that maximized the worst performance of a single criterion among multiple criteria, and addressed potential model misspecification by considering multiple models. One drawback of this approach is that robustness is achieved at the cost of design complexity due to the need to specify multiple criteria and models. Ideally, one would like to achieve reasonable efficiency for assessing hormesis by specifying a single criteria/model.

In the current work we consider the use of Bayesian decision theoretic designs for efficient dose allocation when inference regarding the presence of a hormetic dose-resoponse relationship is of interest. Specifically, we explore two possible choices for the loss function as criteria for

dose allocation. The first is a c-optimal design [16, 22] based upon a multistage model with a modification to include a flexible hormetic zone [1, 20, 37]. In the multistage model, the existence of a hormetic effect is determined by a single parameter. As such, the c-optimal design seeks to maximise statistical information regarding this determinant parameter. The second loss function is based upon Shannon's entropy [55] and is independent of parametric assumptions for the dose-response model. A Bayesian theoretic approach based upon a single loss function provides a simple design specification. Further, implementing the design by specifying a flexible parametric dose-response model (the c-optimal design) or by specifying the loss function to be free of parametric assumptions regarding the model (Shannon's loss) provides robustness to model misspecification. Finally, to account for uncertainty in model specification at the study design stage (before observed data are gathered), we consider a group sequential approach where a series of dose allocation decisions are made periodically as observed responses are obtained.

The remainder of Chapter 6 is organized as follows: Section 6.2 describes a Bayesian decision theoretic approach to the design of hormesis studies under a multistage model. The approach allows for a group sequential allocation scheme in which a series of dose allocation decisions are made periodically with model parameter information updated as observed responses are obtained. Section 6.3 defines and illustrates the use of loss functions based upon a c-optimal design under the multistage model and based upon Shannon's entropy. Section 6.4 presents simulation studies to compare the operating characteristics of the proposed designs and investigates the effect of increasing the number of possible experimental dose points under the proposed methods. Section 6.5 provides a brief discussion of the role of optimal design in risk assessment studies and considers avenues for future research.

6.2 Methods

Bayesian decision theory is a fundamental approach to decision making based on prior knowledge combined with observed data. In the decision theoretic approach, we need four components including an action space (i.e. a set of experimental doses in the current context), a likelihood function, a prior density function and a loss function. Given the four components, a decision (i.e. dose allocation) is made in order to minimize the expected loss.

6.2.1 Action Space

We denote a discrete action space by $\mathcal{D} = \{d_1, \ldots, d_I\}$ with $d_1 < \cdots < d_I$, where $d_1 = 0$ (the control dose group) and $d_I = 1$ (the highest dose group) without loss of generosity. If we need to make k simultaneous decisions, the action space expands to $\mathcal{D}^k = \mathcal{D} \times \cdots \times \mathcal{D}$.

6.2.2 Likelihood Function

We denote n independent Bernoulli random variables by $\vec{Y}_n = (Y_1, \ldots, Y_n)$, where $Y_i = 1$ if the i^{th} experimental unit allocated at dose $x_i \in \mathcal{D}$ shows a prespecified toxic event and $Y_i = 0$, otherwise. The probability of $Y = 1$ associated with dose x is denoted by π_x. If π_x is completely characterized by a vector of model parameters, $\vec{\beta}$, we write $\pi_x(\vec{\beta})$. Then the likelihood function after observing $\vec{Y}_n = \vec{y}_n$ is

$$l(\vec{\beta}) = \prod_{i=1}^{n} \left\{\pi_{x_i}(\vec{\beta})\right\}^{y_i} \left\{1 - \pi_{x_i}(\vec{\beta})\right\}^{1-y_i} . \tag{6.1}$$

We parameterize a dose-response relationship by

$$\pi_x(\vec{\beta}) = 1 - \exp\left\{-\left(\beta_0 + \beta_1 x^{\beta_3} + \beta_2 x^{2\beta_3}\right)\right\} \tag{6.2}$$

for dose $x \geq 0$. This parameterization is modified from the monotonic multistage model using a linear predictor $\sum_{m=0}^{M} \beta_m x^m$ with nonnegative coefficients [20]. As opposed to the linear predictor $\sum_{m=0}^{2} \beta_m x^m$, the non-linear predictor $\sum_{m=0}^{2} \beta_m (x^{\beta_3})^m$ allows for an asymmetric hormetic zone. The parameter space for $\vec{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ is a subset of $(0, \infty) \times (-\infty, \infty) \times (0, \infty) \times (0, \infty)$, with restrictions imposed to guarantee $\pi_x(\vec{\beta}) \in (0, 1)$. In model (6.2), the sign of β_1 determines whether a dose-response relationship is monotonic (i$\beta_1 \geq 0$) or hormetic ($\beta_1 < 0$). Thus the *monotonic parameter space* is

$$\Omega_M = \left\{ \vec{\beta} \in (0, \infty) \times [0, \infty) \times (0, \infty) \times (0, \infty) \right\},$$

and it can be shown that the *hormetic parameter space* is

$$\Omega_H = \left\{ \vec{\beta} \in (0, \infty) \times (-\infty, 0) \times (0, \infty) \times (0, \infty) : \frac{\beta_1^2}{\beta_0 \beta_2} < 4 \right\}.$$

The entire parameter space is the union of the two disjoint parameter spaces,

$$\Omega = \Omega_M \cup \Omega_H = \left\{ \vec{\beta} \in (0, \infty) \times (-\infty, \infty) \times (0, \infty) \times (0, \infty) : 0 < \pi_x(\vec{\beta}) < 1 \right\}.$$

Under model (6.2), the *background risk* is given by $\pi_0(\vec{\beta}) = 1 - e^{-\beta_0}$, and serves as a reference for a hormetic effect at low doses. Specifically, in the presence of hormesis the *hormetic zone* is defined as the set of all doses x such that $\pi_x(\vec{\beta}) < \pi_0(\vec{\beta})$. The *length of hormetic zone* is defined as the value δ such that $\pi_\delta(\vec{\beta}) = \pi_\alpha(\vec{\beta})$ and $\pi_x(\vec{\beta}) < \pi_0(\vec{\beta})$ for all $x \in (0, \delta)$, while the *most beneficial dose* is defined as the dose α which produces the lowest probability of a toxic event. Under model (6.2), the length of the hormetic zone is given by $\delta(\vec{\beta}) = (-\beta_1/\beta_2)^{1/\beta_3}$, and the most benefical dose is given by $\alpha(\vec{\beta}) = (-\beta_1/2\beta_2)^{1/\beta_3}$. Finally, it is common to quantify the *maximal hormetic effect* by the odds ratio comparing the odds of response at the control dose to the odds of response at the most beneficial dose. Thus, letting γ denote

the maximal hormetic effect, based upon model (6.2) we have

$$\gamma(\vec{\beta}) = \left(\frac{\pi_0(\vec{\beta})}{1-\pi_0(\vec{\beta})}\right)\left(\frac{1-\pi_\alpha(\vec{\beta})}{\pi_\alpha(\vec{\beta})}\right).$$

For the remainder of the chapter we simplify notation by suppressing the dependence of α, δ, and γ on $\vec{\beta}$ with the understanding that each parameter is derived from the multistage model given in (6.2).

Finally, it should be noted that the four-parameter multistage model given in (6.2) with the non-linear predictor may be overly flexible in a sense that an estimated $\vec{\beta}$ may belong to Ω_H, despite a monotonic empirical dose-response relationship. To control over flexibility and disregard practically unimportant hormetic effects, we can consider the *restricted parameter space* $\Omega^* = \Omega_M \cup \Omega_H^*$, where

$$\Omega_H^* = \left\{\vec{\beta} \in \Omega_H : \alpha \geq \alpha_L, \delta \geq \delta_L, \gamma \geq \gamma_L\right\}$$

for some $\alpha_L \geq d_2$, $\delta_L \geq d_3$, and $\gamma_L > 1$.

6.2.3 Prior Density Function

Here we consider the specification of prior distributions for the parameters defining the multistage model presented in Section 6.2.2. It is generally difficult to elicit prior knowledge directly through $\vec{\beta}$ from subjet-matter experts. For a tractable prior elicitation, we consider a similar approach to a conditional mean prior [6] which allows specification of the full prior distribution by placing a prior probability distribution on the response at three arbitrary dose levels. More specifically, we choose three arbitrary doses, say $x_{-2} < x_{-1} < x_0$, and specify three independent Beta distributions $\pi_{x_i} \sim \text{Beta}(a_i, b_i)$ for $i = -2, -1, 0$. For mathematical convenience, we fix β_3 then transform the joint probability distributions of $(\pi_{x_{-2}}, \pi_{x_{-1}}, \pi_{x_0})$

to the conditional joint probability distribution of $(\beta_0, \beta_1, \beta_2)$ given β_3. The transformation leads to

$$f(\beta_0, \beta_1, \beta_2 \mid \beta_3) \propto \mathbf{I}_{\vec{\beta} \in \Omega^*} \prod_{i=-2}^{0} \left\{ \pi_{x_i}(\vec{\beta}) \right\}^{a_i - 1} \left\{ 1 - \pi_{x_i}(\vec{\beta}) \right\}^{b_i} , \tag{6.3}$$

where $\mathbf{I}_{\vec{\beta} \in \Omega^*} = 1$ if $\vec{\beta} \in \Omega^*$ and $\mathbf{I}_{\vec{\beta} \in \Omega^*} = 0$ otherwise. Then, we choose a functional form of $f(\beta_3)$ for $\beta_3 > 0$.

If an investigator desires to express informative prior knowledge, it will be important to understand the intepretation of β_3. By denoting $\rho = \alpha/\delta \in (0, 1)$, the ratio ρ determines the degree of asymmetry in the hormetic zone under the model. Then it can be shown that $\beta_3 = \log(0.5) \,/\, \log(\rho)$. Based on this result, $\beta_3 = 1$ implies a symmetric hormetic zone, $\beta_3 = 0.5$ (i.e. $\rho = 0.25$) implies the most beneficial dose α is closer to the control dose and $\beta_3 = 2.41$ (i.e. $\rho = 0.75$) implies α is closer to the end of the hormetic zone δ.

6.2.4 Loss Functions

In this section, we discuss two loss functions. The first loss functions is devised from Bayesian c-optimal designs. A Bayesian c-optimal design seeks precise estimation of a parameter of interest, and it particularly minimizes the expected asymptotic variance of the maximum likelihood estimator under an assumed parametric model [16, 22]. With our current focus on optimal design to infer the presence of hormesis, we are specifically interested in precise estimation of β_1 under the multistage model. The second loss function we present is based upon Shannon entropy and is defined independent of any parametric assumptions regarding the dose-response model.

c-Optimal Design for the Determinant Parameter

The expected Fisher information matrix based on n allocated doses, say $\vec{x}_n = (x_1, \ldots, x_n)$, is denoted by $\mathcal{I}(\vec{x}_n; \vec{\beta})$. Its inverse is denoted by $\mathcal{I}^{-1}(\vec{x}_n; \vec{\beta})$. To maximize information about β_1, the determinant for hormesis, the loss function of the c-optimal design is given by

$$\mathcal{L}_{\beta_1}(\vec{x}_n; \vec{\beta}) = \left\{\mathcal{I}^{-1}(\vec{x}_n; \vec{\beta})\right\}_{22} ,$$

the $(2,2)^{\text{th}}$ element of the inverted matrix. See Appendix A.6 for a more explicit expression. This loss function is referred to hereafter as β_1-*loss* and is simply denoted by $\mathcal{L}_{\beta_1}$.

c-Optimal Design for the Most Beneficial Dose

To maximize information about α, the most beneficial dose in the presence of hormesis (i.e. given $\beta_1 < 0$), the loss function of the c-optimal design is given by

$$\mathcal{L}_{\alpha}(\vec{x}_n; \vec{\beta}) = \vec{g}^T(\vec{\beta})\ \mathcal{I}^{-1}(\vec{x}_n; \vec{\beta})\ \vec{g}(\vec{\beta})$$

with the gradient vector

$$\vec{g}^T(\vec{\beta}) = \frac{\alpha}{\beta_3}\begin{pmatrix} 0 & \beta_1^{-1} & -\beta_2^{-1} & -\log(\alpha) \end{pmatrix} .$$

Shannon-Loss

In contrast to the β_1-loss function that is derived under model (6.2), the following loss function is specified via point-wise assumptions at given experimental doses. For $i = 1, \ldots, I$, we let $\pi^*_{d_i} = \max\{\pi_0, \pi_{d_i}\}$ which is non-decreasing with respect to $d_i \in \mathcal{D}$. Further, we let $m_i = \sum_{j=1}^{n} \mathbf{I}_{x_j = d_i}$, the number of experimental units allocated at $d_i \in \mathcal{D}$ among n

experimental units. Defining $p_i = (m_i + 1)\,\pi^*_{d_i}$ and normalizing by

$$p_i^* = \frac{p_i}{\sum_{i=1}^{I} p_i}\,,$$

$(p_1^*, \ldots, p_I^*)$ forms a legitimate discrete probability distribution on the finite support $\mathcal{D}$. Based on the Shannon's entropy [55], we consider the loss function

$$\mathcal{L}_s(\vec{x}_n) = -\left(-\sum_{i=1}^{I} p_i^* \log(p_i^*)\right) = \sum_{i=1}^{I} p_i^* \log(p_i^*)\,.$$

Because Shannon's entropy is maximized when $p_i^* = 1/I$, use of this loss function will distribute experimental units so that a higher concentration of observations are made inside an estimated hormetic zone relative to outside the zone. Note that the addition of one to each m_i in the definition of p_i avoids an unbounded loss. If we assume a dose-response relationship by the multistage model in Equation (6.2), the loss function $\mathcal{L}_s(\vec{x}_n)$ is denoted by $\mathcal{L}_s(\vec{x}_n; \vec{\beta})$ as $\vec{\beta}$ completely determines $(p_1^*, \ldots, p_I^*)$. This loss function is referred to hereafter as Shannon-loss and is simply denoted by $\mathcal{L}_s$.

6.2.5 Posterior Density Function and Minimizing Average Loss

Let N denote the total sample size available in a dose-response study. Considering model uncertainty and prior uncertainty, allocating all N experimental units at once may not be an optimal strategy. On the other hand, making N truly sequential decisions may not be practical, particularly when observing a toxic event may require a long period of time. As such, we consider making a series of decisions in groups, and we illustrate a two-phase design here. In a two-phase design, we spend n_1 experimental units based on prior distributional assumptions, realize the n_1 independent Bernoulli random variables, spend the remaining $N - n_1 = n_2$ experimental units based on the updated knowledge, realize the n_2 indepen-

dent Bernoulli random variables, then make final inference based on the complete set of N observations together with the prior knowledge.

From Equations (6.2) and (6.3), the joint posterior density function of $\vec{\beta}$ is

$$f(\vec{\beta} \mid \vec{y}_n) \propto \mathbf{I}_{\vec{\beta} \in \Omega^*} f(\beta_3) \prod_{i=-2}^{0} \left\{\pi_{x_i}(\vec{\beta})\right\}^{a_i - 1} \left\{1 - \pi_{x_i}(\vec{\beta})\right\}^{b_i} \prod_{i=1}^{n} \left\{\pi_{x_i}(\vec{\beta})\right\}^{y_i} \left\{1 - \pi_{x_i}(\vec{\beta})\right\}^{1-y_i},$$

where Ω^* is the restricted parameter space given in Section 6.2.2, $f(\beta_3)$ is a prior density function of β_3 and (a_i, b_i) are fixed hyperparameters for the Beta distributions for $i = -2, -1, 0$. For a given loss function, $\mathcal{L}_{\beta_1}$ or $\mathcal{L}_s$, we decide $\vec{x}^*_{n_1} = (x^*_1, \ldots, x^*_{n_1})$ such that the prior average

$$E\left\{\mathcal{L}(\vec{x}^*_{n_1}; \vec{\beta})\right\} = \int \mathcal{L}(\vec{x}^*_{n_1}; \vec{\beta})\, f(\vec{\beta})\, d\vec{\beta}$$

is minimumized with respect to $(x_1, \ldots, x_{n_1}) \in \mathcal{D}^{n_1}$. For given $\vec{x}^*_{n_1}$ and observed $\vec{y}_{n_1}$, we make a second decision $(x^*_{n_1+1}, \ldots, x^*_N)$ such that the posterior average

$$E\left\{\mathcal{L}(\vec{x}^*_N; \vec{\beta}) \,\middle|\, \vec{y}_{n_1}\right\} = \int \mathcal{L}(\vec{x}^*_N; \vec{\beta})\, f(\vec{\beta} \mid \vec{y}_{n_1})\, d\vec{\beta}$$

is minimized with respect to $(x_{n_1+1}, \ldots, x_N) \in \mathcal{D}^{n_2}$. Through the posterior density function

$$f(\vec{\beta} \mid \vec{y}_{n_1}) \propto \mathbf{I}_{\vec{\beta} \in \Omega^*} f(\beta_3) \prod_{i=-2}^{0} \left\{\pi_{x_i}(\vec{\beta})\right\}^{a_i - 1} \left\{1 - \pi_{x_i}(\vec{\beta})\right\}^{b_i} \prod_{i=1}^{n_1} \left\{\pi_{x^*_i}(\vec{\beta})\right\}^{y_i} \left\{1 - \pi_{x^*_i}(\vec{\beta})\right\}^{1-y_i},$$

the observed binary responses $\vec{y}_{n_1} = (y_1, \ldots, y_{n_1})$ at the allocated doses $\vec{x}^*_{n_1} = (x^*_1, \ldots, x^*_{n_1})$, together with the prior knowledge expressed at (x_{-2}, x_{-1}, x_0), contribute to the second decision $(x^*_{n_1+1}, \ldots, x^*_N)$. After observing all responses $\vec{y}_N$ at $\vec{x}^*_N = (x^*_1, \ldots, x^*_N)$, we make the final inference based on $f(\vec{\beta} \mid \vec{y}_N)$.

6.3 Illustrations

In this section, we describe the distinguishable characteristics of β_1-loss and Shannon-loss, with respect to the distribution of experimental units using a hypothetical example. For a conditional mean prior, we select $x_{-2} = 0$, $x_{-1} = 0.2$, and $x_0 = 0.5$ and specify $\pi_{x_{-2}} \sim$ Beta$(2.6, 7.5)$, $\pi_{x_{-1}} \sim$ Beta$(1.5, 5.4)$ and $\pi_{x_0} \sim$ Beta$(1.5, 1.5)$. This is a prior that weakly favours hormesis over monotonicity. A flat prior is used for β_3. For an action space $\mathcal{D}$, we consider $I = 11$ evenly spaced experimental doses such that $\mathcal{D} = \left\{0, \frac{1}{10}, \ldots, \frac{9}{10}, 1\right\}$. We assume that the true dose-response relationship is given by the four-parameter multistage model with $\vec{\beta} = (0.22, -1.66, 4.15, 1)$. This yields a background risk of $\pi_0 = 0.2$, most beneficial dose of $\alpha = 0.2$ with the minimum risk $\pi_\alpha = 0.05$, length of hormetic zone of $\delta = 0.4$, and maximal hormetic effect of $\gamma = 4.26$. For the restricted parameter space $\Omega^* = \Omega_M \cup \Omega_H^*$, we set $\alpha_L = d_2 = 0.1$, $\delta_L = d_3 = 0.2$ and $\gamma_L = 1.1$. Using the two-phase design with $n_1 = n_2 = 100$ (i.e. $N = 200$), we apply the two decision theoretic approaches with the loss functions $\mathcal{L}_{\beta_1}$ and $\mathcal{L}_s$. For the purpose of illustration, we artificially generate the most likely responses at each phase such that the observed number of toxic events at d_i is the nearest integer to $m_i \times \pi_{d_i}$.

We obtain the results shown in Figure 6.1 by using $\mathcal{L}_{\beta_1}$ and Figure 6.3 by using $\mathcal{L}_s$. In each figure, the dose allocations are marked by "o", and the size of the symbol is relative to the maximum of $(m_1, \ldots, m_{11})$. The number of experimental doses at each unit is provided at the top of each figure. In addition, two estimated dose-response curves are drawn conditioning on $\vec{\beta} \in \Omega_M$ and $\vec{\beta} \in \Omega_H^*$. The fitted curve under $\vec{\beta} \in \Omega_H^*$ appears to be nearly perfect because the most likely responses are generated under the model. As shown in Figure 6.1, the resulting allocations are (86, 1, 25, 38, 0, 0, 0, 9, 11, 26, 3) using $\mathcal{L}_{\beta_1}$, and the posterior probability of hormesis is $P(\Omega_H^* \mid \vec{y}_{200}) = 0.998$. Three quarters of the experimental units are allocated inside the true hormetic zone $(0, 0.4)$. As shown in Figure 6.2, allocated doses using $\mathcal{L}_\alpha$ are concentrated at the four doses (0, 0.3, 0.5, 1) with the respective group sizes

(77, 47, 48, 26). Under the model assumption with $\beta \in \Omega_H^*$, the loss function achieves a precise estimate of α by allocation around the estimated α rather than inside the estimated hormetic zone. To this end, less than two thirds of the experimental units are allocated inside the hormetic zone, and the posterior probability of hormesis is 0.967. Figure 6.3 shows that $\mathcal{L}_s$ results in (36, 36, 36, 35, 27, 12, 6, 4, 3, 3, 2), and the posterior probability of hormesis is 0.969. Approximately 85% of experimental units are allocated inside the hormetic zone as the Shannon-loss function seeks to obtain relatively more information inside the estimated hormetic zone.

Figure 6.1: Example of the $\mathcal{L}_{\beta_1}$-Design.

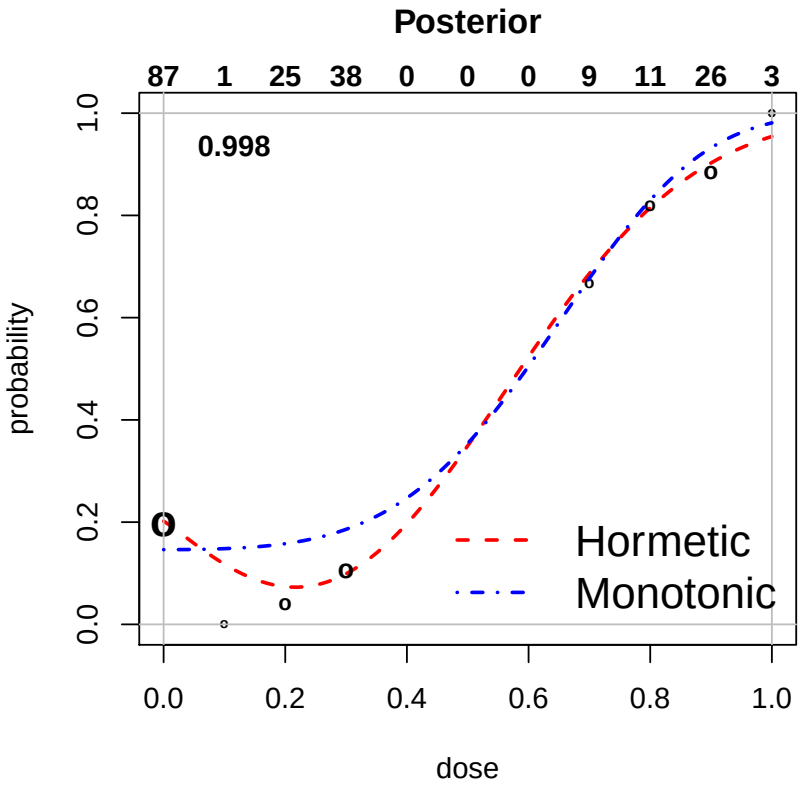
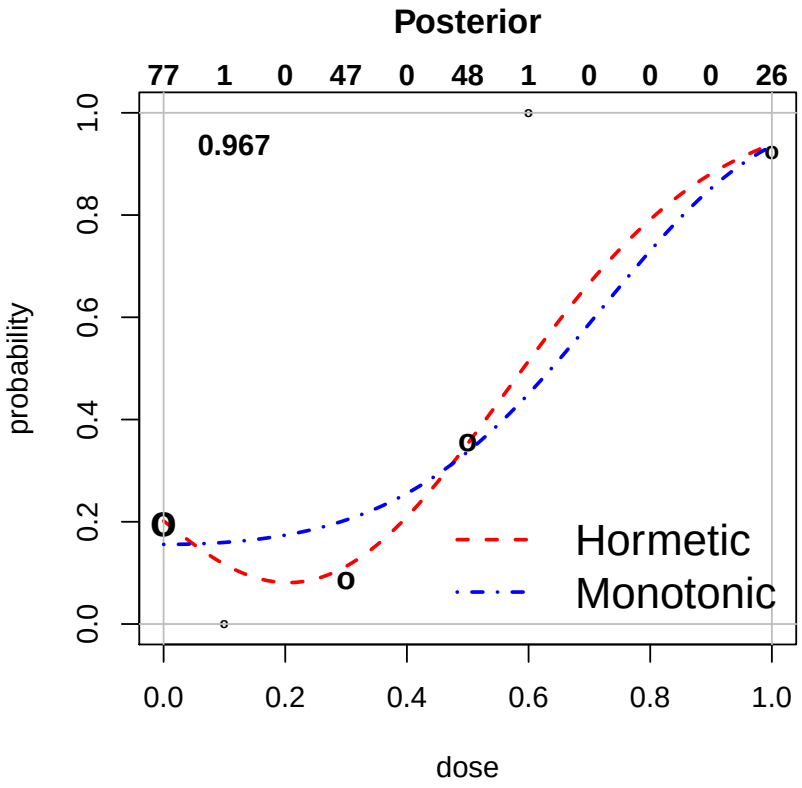


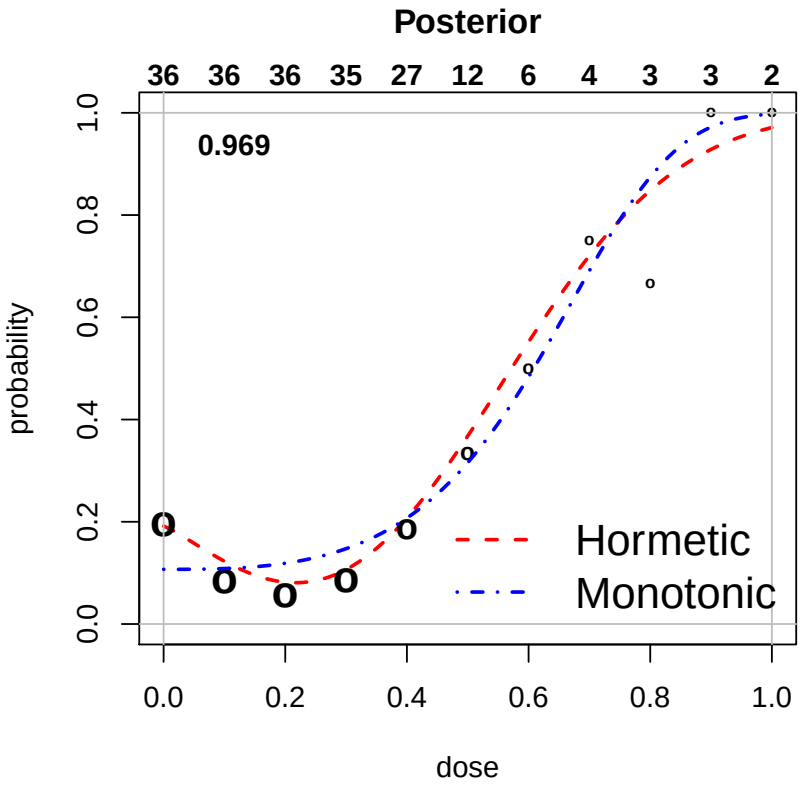
Figure 6.2: Example of the $\mathcal{L}_\alpha$-Design



6.4 Simulations

We conducted simulation studies were conducted to compare four different designs with respect to estimation of δ and sensitivity and specificity for testing the hypothesis of a hormetic dose-response relationship based upon the posterior probability of hormesis. The first two designs are the two-stage designs based on the decision theoretic approaches using β_1- and Shannon-loss. Both designs are implemented under the under the same action space ($I = 11$ evenly spaced experimental doses such that $\mathcal{D} = \left\{0, \frac{1}{10}, \ldots, \frac{9}{10}, 1\right\}$), the same prior, the same restrictions on the hormetic parameter space and $n_1 = n_2 = 100$ as in Section 3. The remaining two designs are balanced designs with $I = 5$ and $I = 11$ (i.e. the same number of experimental units at each experimental units). We denote the designs as $\mathcal{B}_5$ and $\mathcal{B}_{11}$, respectively. For the total sample size of $N = 200$, $\mathcal{B}_5$ a priori fixes 40 experimental

Figure 6.3: Example of the $\mathcal{L}_s$-Design



units on $\mathcal{D} = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}$ and $\mathcal{B}_5$ a priori fixes $m_1 = m_2 = 19$ and $m_3 = \cdots = m_{11} = 18$. $\mathcal{B}_5$ reflects a commonly implemented study design in risk assessment studies. $\mathcal{B}_{11}$ is included in order to compare a balanced approach to the Bayesian decision theoretic designs under the same action space.

6.4.1 Scenarios

Under the four-parameter multistage model, we generated Scenario 0 with $\vec{\beta} = (0.22, 0.00, 2.00, 2.00)$. The resulting dose-response curve is monotonic (i.e. $\vec{\beta} \in \Omega_M$) and close to a threshold dose-response relationship as $\pi_0(\vec{\beta}) = 0.197$, $\pi_{0.2}(\vec{\beta}) = 0.200$, $\pi_{0.5}(\vec{\beta}) = 0.292$ and $\pi_1(\vec{\beta}) = 0.891$. We then generated hormetic Scenarios 1 to 9 with various $\vec{\beta} \in \Omega_H^*$ as shown

in Figure 6.4. Scenarios 1 to 3 are representative cases for relatively long hormetic zones (see Figure 6.4a), Scenarios 4 to 6 represent hormetic cases with moderate lengths of hormetic zones (see Figure 6.4b), and Scenarios 7 to 9 represent hormetic cases with relatively short hormetic zones (see Figure 6.4c). From Scenario 1 to Scenario 9, the length of hormetic zone is non-increasing. Simulation results were based upon 1000 simulated studies for each scenario.

6.4.2 Evaluation Methods

In each simulated trial of a hormetic case, we estimate α and δ. We evaluate the quality of point estimation by bias, standard error, and root mean square error (RMSE) in percent. For example, the percent bias in the estimation of δ is bias / $\alpha_0 \times 100\%$, where δ_0 is the true value of δ in each hormetic scenario. We also evaluate the coverage probability of 95% credible intervals (CIs) by calculating the proportion of 95% CIs including the true parameter value.

In each simulated trial, we conclude whether the dose-response relationship is monotonic or hormetic based on $P(\Omega_H^* \mid \vec{y}_{200})$. Hormesis is concluded if $P(\Omega_H^* \mid \vec{y}_{200}) \geq q$ for some threshold $q \in (0, 1)$, and monotonicity is concluded otherwise. By repeating each scenario a large number of times, we can estimate the area under a receiver operating characteristic (ROC) curve in the context of hypothesis testing. If we let Π_q denote the probability of concluding hormesis for a fixed decision threshold q, Π_q is one minus specificity in Scenario 0, and it is sensitivity in Scenarios 1 to 9. Using $P(\Omega_H^* \mid \vec{y}_{200})$ in each simulated trial, we can estimate Π_q for all $q \in (0, 1)$ under each scenario. For each hormetic scenario, by varying q from zero to one, we can plot Π_q obtained in Scenario 0 on the x-axis and Π_q obtained in the hormetic scenario on the y-axis. Then we obtain the ROC curve and estimate the area under the ROC curve (AUR) to quantify the performance of a design with respect to sensitivity and specificity in hypothesis testing. An AUR close to 1 reflects a study design with nearly

perfect discrimination between a hormetic and monotonic dose response relationship.

6.4.3 Results

For Scenarios 1 to 3, when the hormetic zones are relatively long, all four designs perform fairly well with respect to parameter estimation and hypothesis testing. As shown in Table 6.1, all designs yield AURs greater than 0.9, except for the $\mathcal{L}_{\beta_1}$-design in Scenario 3 which results in an AUR of 0.830. With respect to percent RMSE, the $\mathcal{B}_5$-design shows the best performance in estimation of δ, while all designs show satisfying coverage probabilities close to 0.95. Scenarios 1 to 3 illustrate that the operating characteristics of each design are not appreciably different when many experimental doses are located inside the true hormetic zone. In particular, the $\mathcal{B}_5$-design, which resembles many past experiments, performs as the Bayesian decision theoretic approaches or a balanced design with a larger action space.

We now focus on Scenarios 4 to 6, where approximately half of the experimental doses in the action space lie within the true hormetic zone. In these cases, the $\mathcal{L}_{\beta_1}$-design performs relatively well, particularly with respect to hypothesis testing. The resulting AURs for the $\mathcal{L}_{\beta_1}$-design are consistently greater than the other designs. The $\mathcal{B}_5$-design reveals deficiencies in these cases, particularly in Scenarios 5 and 6 as none of one thousand resulting 95% CIs for δ covers the true values and the estimated AUR is close to 0.5 in Scenario 6. The $\mathcal{L}_s$-design consistently yields larger AURs than $\mathcal{B}_{11}$-design by allocating more experimental units inside the hormetic zone. The resulting RMSEs are the smallest in the $\mathcal{L}_\alpha$-design while its resulting AURs are relatively small.

We turn our focus to Scenarios 7 to 9 where, in general, the differences between the four designs that were revealed in Scenarios 4 to 6 are amplified in the presence of a shorter hermetic zone. The $\mathcal{L}_{\beta_1}$-design yields the greatest AURs in Scenarios 7 and 8 among the four designs, but it performs worse than Shannon-design in Scenario 9. In these three

scenarios, the respective resulting AUR is 0.605, 0.743 and 0.306 for the $\mathcal{L}_{\beta_1}$-design and 0.539, 0.703 and 0.429 for the $\mathcal{L}_s$-design. Though the $\mathcal{L}_s$ design does not perform the best in all scenarios, it performs well relative to the best design in each scenario with respect to AURs. The resulting large positive biases for δ are potentially due to the two lower bounds $\alpha_L = 0.1$ and $\delta_L = 0.2$ in the restricted parameter space Ω_H^*.

A summary of the resulting AURs across all nine scenarios is presented in Figure 6.5. The figure illustrates that the $\mathcal{L}_{\beta_1}$-design performs well when the hormetic zone is not too short, while the $\mathcal{L}_s$-design tends to be robust with respect to hypothesis testing across the various hormetic zone lengths. In particular, the resulting AUR associated with Shannon-loss is consistently greater than that of the $\mathcal{B}_{11}$-design. This should be an anticipated result, since transferring some experimental units to doses within the hormetic zone increases the chance of detecting hormesis. Finally, the resulting AUR of the $\mathcal{B}_5$-design decreases at the fastest rate as the length of the hormetic zone decreases. This result is particularly concerning as it is the primary design that has been utilised in past studies where the presence of hormesis has been debated [71].

6.4.4 The Effect of the Number of Experimental Doses

Regardless of relative performance, a resulting AUR below 0.5 is undesirable. This occurs for all of the designs under Scenario 9 because the hormetic zone occupies only 20% of the experimental range and $d_2 = \frac{1}{10}$ is the only experimental dose with a hormetic effect. Under this scenario, we investigate the effect of increasing the number of experimental doses, I, and the sample size, N. We focus on the $\mathcal{L}_s$-design with two stages, $N = 200, 300, 400, 500$ ($n_1 = n_2 = N/2$), and $I = 5, 11, 22, 33$, where the experimental doses are evenly spaced in the standardised interval [0,1].

The resulting AURs using the four levels of I and the four levels of N are shown in Table 6.2.

The resulting AUR is 0.82 with $I = 22$ and $N = 300$ which is comparable to the resulting AUR, 0.78, with $I = 11$ and $N = 500$. Surprisingly, the resulting AUR is 0.69 with $I = 22$ and $N = 200$, which is slightly greater than the resulting AUR of 0.65 with $I = 11$ and $N = 400$, despite having twice the maximal sample size. For $N = 200, 300, 400, 500$, $I = 33$ does not yield a larger AUR (0.68, 0.82, 0.89, 0.94, respectively) than $I = 22$ of the same N (0.69, 0.82, 0.92, 0.96, respectively). Increasing N for $I = 5$ is not worthy as expected.

6.5 Discussions

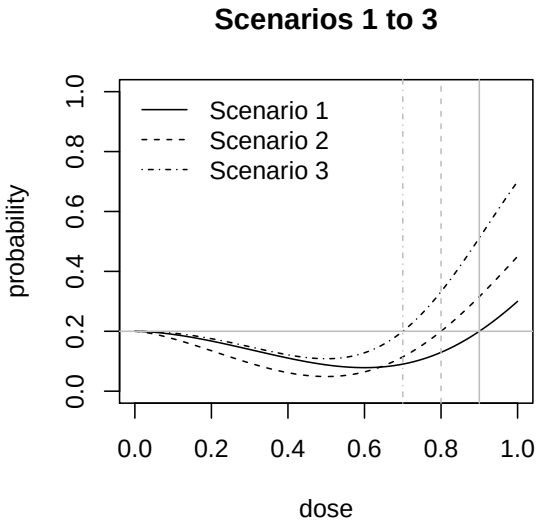
We discussed the operating characteristics of the two Bayesian designs and the traditional balanced designs, and we raised the issue of using the past designs particularly when the number of experimental doses in the true hormetic zone is scarce. The c-optimal design, $\mathcal{L}_{\beta_1}$, was shown to be beneficial when two or more fixed experimental doses are inside the hormetic zone. Overall, the Shannon-design, $\mathcal{L}_s$, demonstrated a relatively robust performance across all considered scenarios. When many experimental doses lie inside the hormetic zone, it was seen that the results are not sensitive to the choice of design. On the other hand, the disadvantage of an inefficient design is revealed as the number of experimental doses inside the hormetic zone decreases.

Though we have not presented it in this chapter, we also investigated another c-optimal design seeking precise estimation of α, the most beneficial dose, under the assumption of a hormetic effect (i.e. $\beta_1 < 0$). Interestingly, a relatively large number of experimental units are allocated around the hormetic zone rather than inside the hormetic zone. Under a correctly specified model together with the existence of hormesis, this is an efficient method to reduce the variability in the estimate of the most beneficial does but it did not show satisfying results with respect to the AUR for discriminating between a monotonic or hormetic dose-response relationship.

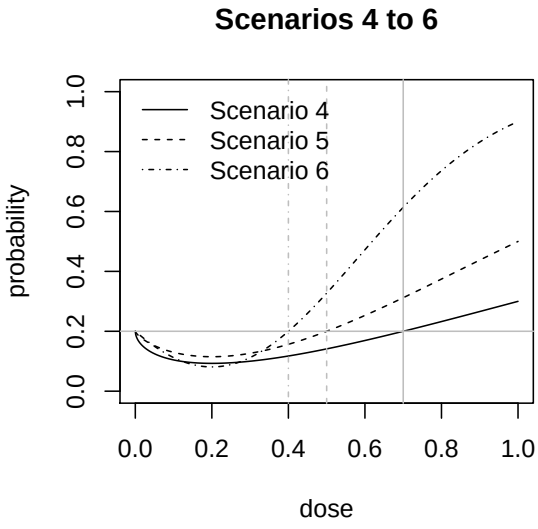
One advantage of Shannon-loss is the applicability to both parametric and non-parametric models. Furthermore, since it does not target a specific objective under a specified model, it generally shows robustness with respect to multiple objectives for risk assessment at low dose. Further investigations and improvements remain necessary since a resulting AUR less than 0.5 is not desirable in any case. One possible approach is to modify the loss function so that more experimental units are allocated to the control dose. Currently we weigh the importance of the control dose equal to the importance of any experimental dose inside an estimated hormetic zone. However, a more accurate estimate of the background risk may be desirable because the definition of a hormetic zone (and its existence) is relative to the risk at the control dose. Another possible approach is to vary the action space at the time of a new decision. In our numerical experiments, it appears that the relative length of the true hormetic zone to the experimental range is a crucial factor for detecting hormesis. By reducing the range of the action space based on updated knowledge, we may be able to reduce unnecessary allocations at high doses. Of course, these modifications would add to the complexity of implementing the designs and that tradeoff would need to be considered in practice.

The extended simulation study of Scenario 9 in Section 4.4 shows that increasing the size of the action space, I, achieves higher efficiency with the same or less N. In particular, we demonstrated one case where doubling the density of the action space for for a given sample size is more efficient than doubling the sample size for the same density. Kavlock [36] addressed a similar point with respect to the estimation of a benchmark dose [26]. The take-home message from this result is that dramatic savings in sample size can be obtained by adding fairly minimal logistical effort, a finding that has broad scientific and ethical implications in animal-based studies.

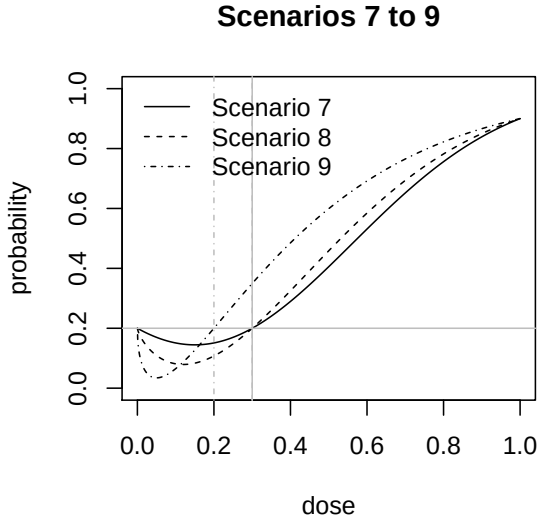
Figure 6.4: Simulation scenarios



(a) Scenarios 1 to 3



(b) Scenarios 4 to 6



(c) Scenarios 7 to 9

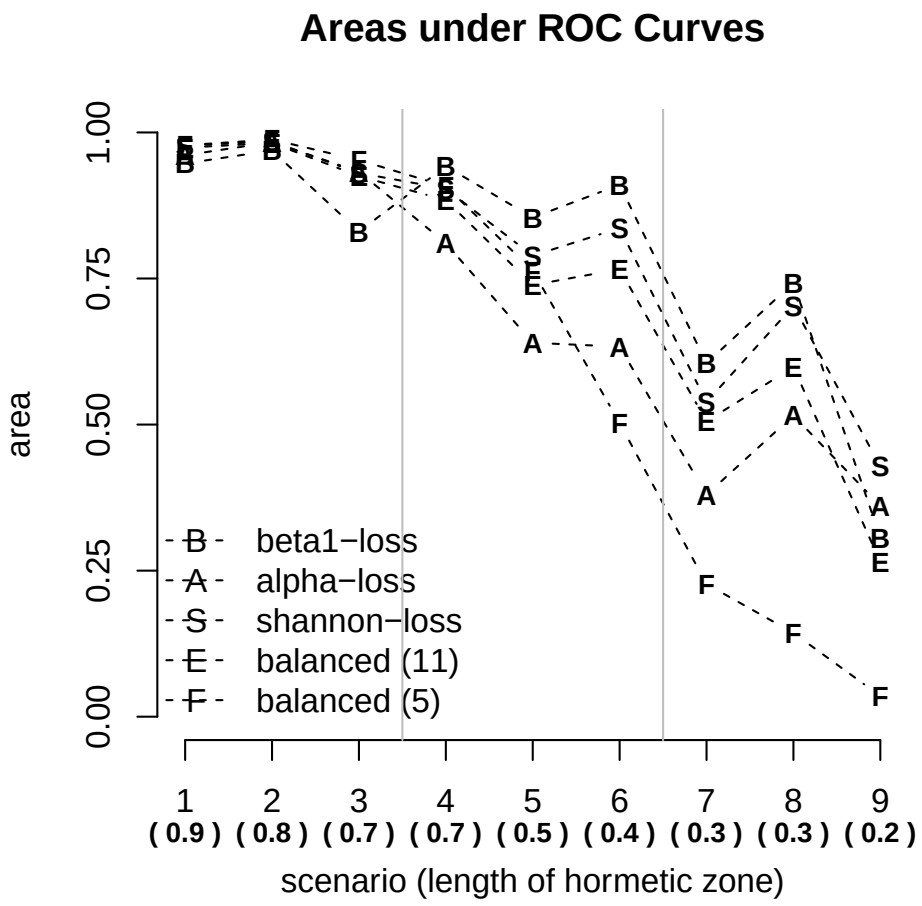
Table 6.1: Simulation results

		Estimation of α				Estimation of δ				
Scenario	**Design**	Bias (%)	SE (%)	RMSE (%)	Coverage	Bias (%)	SE (%)	RMSE (%)	Coverage	AUR
1	$\mathcal{L}_{\beta_1}$	-33.7	20.4	39.4	0.859	-0.3	23.8	23.8	0.959	0.947
	$\mathcal{L}_{\alpha}$	-21.4	17.7	27.7	0.911	20.2	48.3	52.4	0.962	0.961
	$\mathcal{L}_{s}$	-19.0	16.9	25.5	0.932	18.0	40.4	44.2	0.953	0.977
	$\mathcal{B}_{11}$	-19.0	18.2	26.5	0.920	20.1	41.8	46.4	0.948	0.974
	$\mathcal{B}_{5}$	-16.6	13.7	21.6	0.930	6.1	18.2	19.1	0.965	0.978
2	$\mathcal{L}_{\beta_1}$	-30.2	20.0	36.3	0.852	-9.3	14.5	17.3	0.920	0.969
	$\mathcal{L}_{\alpha}$	-11.4	13.6	17.9	0.930	5.7	16.5	17.3	0.978	0.983
	$\mathcal{L}_{s}$	-11.6	14.6	18.4	0.936	3.4	14.0	14.6	0.968	0.985
	$\mathcal{B}_{11}$	-13.3	15.0	20.0	0.927	2.1	12.0	12.2	0.967	0.982
	$\mathcal{B}_{5}$	-10.0	12.7	16.1	0.956	0.1	9.0	9.4	0.975	0.988
3	$\mathcal{L}_{\beta_1}$	-35.2	21.6	41.2	0.823	-17.7	15.5	23.6	0.902	0.830
	$\mathcal{L}_{\alpha}$	-17.1	15.6	23.2	0.852	-0.4	11.6	11.3	0.969	0.933
	$\mathcal{L}_{s}$	-17.2	16.5	23.7	0.866	-1.0	11.7	12.0	0.973	0.931
	$\mathcal{B}_{11}$	-15.3	15.4	21.9	0.867	-2.1	10.0	10.0	0.980	0.925
	$\mathcal{B}_{5}$	-11.2	13.3	17.3	0.891	-1.7	8.2	8.5	0.981	0.953
4	$\mathcal{L}_{\beta_1}$	29.4	33.7	44.7	0.999	12.4	34.0	36.3	0.961	0.942
	$\mathcal{L}_{\alpha}$	18.7	35.5	40.0	0.982	20.9	45.5	50.0	0.978	0.810
	$\mathcal{L}_{s}$	25.7	36.0	44.2	0.985	16.9	43.3	46.4	0.949	0.904
	$\mathcal{B}_{11}$	23.3	37.2	44.2	0.994	19.2	36.1	40.9	0.968	0.884
	$\mathcal{B}_{5}$	75.3	23.5	79.1	0.000	21.8	24.5	32.7	0.917	0.908
5	$\mathcal{L}_{\beta_1}$	12.4	27.8	30.0	0.997	9.9	19.3	21.4	0.982	0.854
	$\mathcal{L}_{\alpha}$	7.4	27.0	28.3	0.980	20.5	24.5	31.9	0.985	0.640
	$\mathcal{L}_{s}$	12.4	30.5	33.2	0.984	11.6	21.9	24.9	0.966	0.789
	$\mathcal{B}_{11}$	5.1	26.9	27.4	0.997	13.5	20.4	24.5	0.972	0.739
	$\mathcal{B}_{5}$	58.9	16.3	61.2	0.000	30.1	12.1	32.6	0.001	0.764
6	$\mathcal{L}_{\beta_1}$	9.3	24.5	26.5	0.990	4.5	13.9	14.1	0.991	0.910
	$\mathcal{L}_{\alpha}$	1.7	17.0	17.3	0.968	4.8	13.5	14.1	0.987	0.633
	$\mathcal{L}_{s}$	1.4	20.2	20.0	0.961	0.5	13.2	13.2	0.972	0.836
	$\mathcal{B}_{11}$	0.3	20.7	21.2	0.973	2.0	14.3	14.1	0.989	0.766
	$\mathcal{B}_{5}$	50.5	12.0	52.0	0.000	32.5	3.0	32.8	0.000	0.503
7	$\mathcal{L}_{\beta_1}$	25.9	33.7	42.4	0.981	20.0	19.9	28.3	0.944	0.605
	$\mathcal{L}_{\alpha}$	13.6	22.1	25.8	0.949	22.8	17.9	28.9	0.960	0.380
	$\mathcal{L}_{s}$	18.1	24.8	30.6	0.968	17.8	16.7	24.5	0.961	0.539
	$\mathcal{B}_{11}$	14.7	25.4	29.4	0.976	21.6	17.3	27.7	0.961	0.506
	$\mathcal{B}_{5}$	106.4	18.8	108.0	0.000	76.1	3.3	76.2	0.000	0.227
8	$\mathcal{L}_{\beta_1}$	46.1	33.9	57.0	0.943	17.2	17.0	23.8	0.951	0.743
	$\mathcal{L}_{\alpha}$	14.7	15.5	20.4	0.973	10.8	14.1	17.3	0.983	0.516
	$\mathcal{L}_{s}$	23.0	21.1	31.6	0.945	9.1	14.1	16.3	0.975	0.703
	$\mathcal{B}_{11}$	20.4	21.0	28.9	0.960	13.9	14.6	20.0	0.979	0.598
	$\mathcal{B}_{5}$	146.5	19.7	147.8	0.000	75.0	2.9	75.1	0.000	0.143
9	$\mathcal{L}_{\beta_1}$	179.3	56.6	188.1	0.000	43.4	19.5	47.4	0.000	0.306
	$\mathcal{L}_{\alpha}$	124.4	11.3	124.9	0.000	50.3	17.3	52.9	0.000	0.361
	$\mathcal{L}_{s}$	131.9	18.0	133.4	0.000	37.2	14.3	40.0	0.000	0.429
	$\mathcal{B}_{11}$	139.7	24.4	141.4	0.000	50.7	19.2	54.3	0.000	0.265
	$\mathcal{B}_{5}$	500.9	69.9	505.6	0.000	160.9	3.3	160.9	0.000	0.035

Table 6.2: Effects of the number of experimental doses and sample size

	$N = 200$	$N = 300$	$N = 400$	$N = 500$
$I = 5$	0.0070	0.0145	0.0176	0.0216
$I = 11$	0.4324	0.5511	0.6500	0.7782
$I = 22$	0.6906	0.8187	0.9235	0.9574
$I = 33$	0.6833	0.8184	0.8937	0.9391

Figure 6.5: Resulting areas under ROC curves



Chapter 7

Summaries and Future Directions

7.1 Contributions to the Statistical Literature

7.1.1 Consensus Estimators

Often, multiple scientists are available for a scientific problem, and it is reasonable to incorporate multiple prior opinions to gain robustness. By comparing the two classes of consensus estimators, we illustrated that the weighted posterior estimators are preferable when the true parameter value is located well inside the multiple prior guesses while the consensus prior estimators are preferable when the true parameter value is substantially closer to one prior guess. It is because a consensus prior estimator allows updating the weight of each individual posterior based on observed data, and it decreases bias while it increases variance. Chapter 2 contributes to broad scientific communities that the data-dependent weighting scheme can be well justified in Bayesian framework.

7.1.2 Balancing Individual- and Population-Level Ethics

In small-sample dose-finding studies, investigators may face a dilemma between individual- and population-level ethics, and they shall consider the dilemma. The balancing point will be debatable, and the degree of weighting each level of ethics shall depend on characteristics of a new experimental treatment being investigated. In Chapter 3 we decomposed the loss function which has been known to reflect population-level ethics discovered that it indeed balances individual- and population-level ethics in a long-term. By modifying the loss function with an adjustable parameter, trialists will be able to achieve a balance in small-sample Phase I clinical trials.

7.1.3 Estimation of Benchmark Dose

In estimation of a benchmark dose, it has been common to rely on monotonic dose-response models only. Various methods of averaging multiple monotonic dose-response models have been widely discussed and practiced, and we added an alternative class of assumptions under the multistage model. This idea can be extended to any generalized linear model equipped in the current software in EPA. The main contribution of Chapter 4 is the illustration of robustness gained by allowing both monotonicity and hormesis as possibilities in the estimation of a benchmark dose.

7.1.4 Hypothesis Testing for Hormesis

In hypothesis testing for hormesis, we compared the parametric and nonparametric approaches by evaluating sensitivity and specificity through the receiver operating characteristic curve. For the parametric approach, we considered the non-linear predictor which allows flexible shapes of a hormetic zone. Compared to the existing Hunt-Bowman model, the new

parameterization outperforms when the true hormetic zone is heavily asymmetric. Overall, the nonparametric approach based on Bayesian model averaging shows robustness regardless of the shape and the length of a hormetic zone. Chapter 5 enlightens potential issues of parametric modeling in the hypothesis testing including the effect of leverage in regression-based modeling.

7.1.5 Experimental Designs for Detecting Hormesis

A satisfying result in hypothesis testing for hormesis cannot be achieved by flexible dose-response models only but it also requires careful experimental designs. Based on Bayesian decision theoretic approaches, we proposed the loss function for general purposes in toxicologic studies for cancer risk assessment focusing on low doses. It does not require a parametric model as opposed to the Bayesian optimal designs. In practice, collected data are used for multiple objectives, and this loss function will be suitable to this respect. In addition, using a receiver operating characteristic curve, we illustrated the significant effect of adding more experimental doses inside a hormetic zone. The results presented in Chapter 6 shall motivate practitioners to carefully think about the importance of experimental designs.

7.2 Future Directions

7.2.1 BIGM for Late-Onset Toxicities

In Phase I clinical trials for investigating late-onset toxicities, it is possible that a new patient enters the study before the complete observations of previous patients. Cheung and Chappell [17] accounted for incomplete observations (i.e. toxic events have not observed from previous patients before a new patient enters the study), and we can extend the BIGM

to this situation as well. It is not discussed in this book, but the operating characteristics of the BIGM are well preserved when we account for incomplete observations. We need more investigations through various simulation scenarios including misspecification of a survival model component in the likelihood function.

7.2.2 Semi-Parametric Models for Hormesis

In hypothesis testing for hormesis, it is possible to consider a semi-parametric approach. In the setting of Phase I clinical trials, Yin and Yuan [70] specified multiple semi-parametric models and applied Bayesian model averaging to provide robustness in the estimation of s MTD. They considered a set of monotonic dose-response models, which is a common assumption in Phase I clinical trials. We can extend the semi-parametric approach by combining both monotonic and hormetic models in a model space.

7.2.3 Experimental Designs for Detecting Hormesis

In experimental designs for detecting hormesis, we already have learned that it is important to allocate a relatively large number of experimental units inside the hormetic zone. On the other hand, one possible reason that the L_{β_1}-design outperforms in many scenarios, particularly when a hormetic zone is long, is that it allocates a relatively large number of experimental units at the control group. It is somewhat intuitive because the definition of hormesis is relative to the background risk. It may imply that we can improve the sensitivity and the specificity of the $\mathcal{L}_s$-design by allocating more experimental units to the control dose group rather than the uniform distribution inside an estimated hormetic zone.

Bibliography

[1] P. Armitage. Multistage models of carcinogenesis. *Environmental Health Perspectives*, 63:195–201, 1985.

[2] P. Armitage and R. Doll. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *British Journal of Cancer*, 11:161–169, 1957.

[3] J. Babb, A. Rogatko, A. Rogatko, and S. Zacks. Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Statistics in Medicine*, 17:1103–1120, 1998.

[4] A. J. Bailer and M. W. Wheeler. Model uncertainty and risk estimation for experimental studies of quantal responses. *Risk Analysis*, 25:291–299, 2005.

[5] J. Bartroff and T. L. Lai. Incorporating individual and collective ethics into phase I cancer trial designs. *Biometrics*, 67(2):596–603, 2011.

[6] E. J. Bedrick, R. Christensen, and W. Johnson. A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91:1450–1460, 1996.

[7] R. G. Belz and H. Piepho. Modeling effective dosages in hormetic dose-response studies. *PLoS ONE*, 7(3):e33432. doi:10.1371/journal.pone.0033432, 2012.

[8] K. T. Bogen. Generic hockey-stick model for estimating benchmark dose and potency: Performance relative to bmds and application to anthraquinone. *Dose-response*, 9:182–208, 2011.

[9] S. T. Buckland, K. P. Burnham, and N. H. Augustin. Model selection: An integral part of inference. *Biometrics*, 53:603–618, 1997.

[10] E. J. Calabrese. The future of hormesis: where do we go from here? *Critical Reviews in Toxicology*, 31:637–648, 2001.

[11] E. J. Calabrese and L. A. Baldwin. A quantitatively-based methodology for the evaluation of chemical hormesis. *Human and Ecological Risk Assessment: An International Journal*, 3:545–554, 1997.

[12] E. J. Calabrese and L. A. Baldwin. The frequency of u-shaped dose responses in the toxicological literature. *Unknown*, 62:330–338, 2001.

[13] E. J. Calabrese and L. A. Baldwin. Defining hormesis. *Human and Experimental Toxicology*, 21(2):91–97, 2002.

[14] E. J. Calabrese and L. A. Baldwin. The hormetic dose-response model is more common than the threshold model in toxicology. *Toxicological Sciences*, 71:246–250, 2003.

[15] E. J. Calabrese and R. R. Cook. Hormesis: how it could affect the risk assessment process. *Human and Experimental Toxicology*, 24:265–270, 2005.

[16] K. Chaloner and K. Larntz. Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208, 1989.

[17] Y. K. Cheung and C. R. Sequential designs for phase i clinical trials with late-onset toxicities. *Biometrics*, 56:1177–1182, 2000.

[18] R. T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5:559–583, 1989.

[19] R. R. Cook and E. J. Calabrese. The importance of hormesis to public health. *Environmental Health Perspectives*, 114:1631–1635, 2006.

[20] K. S. Crump. The linearized multistage model and the future of quantitative risk assessment. *Human and Experimental Toxicology*, 15:787–798, 1996.

[21] K. S. Crump. The regulatory implications of hormesis: Is hormesis a universal phenomenon? *Critical Reviews in Toxicology*, 31:669–679, 2001.

[22] H. Dette. A note on bayesian c- and d-optimal designs in nonlinear regression models. *The Annals of Statistics*, 24:1225–1234, 1996.

[23] H. Dette, A. Pepelyshev, and W. K. Wong. Optimal experimental design strategies for detecting hormesis. *Risk Analysis*, 31(12):1949–1960, 2011.

[24] S. D. Durham and N. Flournoy. Random walks for quantile estimation. *Statistical Decision Theory and Related Topics*, V:467–476, 1994.

[25] S. D. Durham, N. Flournoy, and W. F. Rosenberger. A random walk rule for phase I clinical trials. *Biometrics*, 53:745–760, 1997.

[26] EPA. Benchmark dose technical guidance, epa/100/r-12/001. Technical report, Environmental Protection Agency, Risk Assessment Forum, Washington, DC, 2012.

[27] D. Faries. Practical modifications of the continual reassessment method for phase I cancer clinical trials. *Journal of Biopharmaceutical Statistics*, 4, 1994.

[28] N. Flournoy, S. D. Durham, and W. F. Rosenberger. Toxicity in sequential dose-response experiments. *Sequential Analysis*, 14:217–227, 1995.

[29] S. N. Goodman, M. L. Zahurak, and S. Piantadosi. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine*, 14:1149–1161, 1995.

[30] C. N. Haas, J. B. Rose, and C. P. Gerba. *Quantitative Microbial Risk Assessment*. New York: Wiley, 1999.

[31] L. M. Haines, I. Perevozskaya, and W. F. Rosenberger. Bayesian optimal designs for phase I clinical trials. *Biometrics*, 59(3):591–600, 2003.

[32] C. Hans and B. D. Dunson. Bayesian inferences on umbrella orderings. *Biometrics*, 61:1018–1026, 2005.

[33] D. Hunt and S. N. Rai. Testing threshold and hormesis in a random effects dose-response model applied to developmental toxicity data. *Biometrical Journal*, 47(3):319–328, 2005.

[34] D. L. Hunt and D. Bowman. A parametric model for detecting hormetic effects in developmental toxicity studies. *Risk Analysis*, 24(1):65–72, 2004.

[35] H. Ishwaran and J. S. Rao. Spike and slab variable selection: Frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–773, 2005.

[36] R. H. Kavlock, J. E. Schmid, and R. W. Setzer. A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. *Risk Analysis*, 16:391–403, 1996.

[37] S. B. Kim, S. M. Bartell, and D. L. Gillen. Estimation of a benchmark dose in the presence or absence of hormesis using posterior averaging. *Risk Analysis*, in press.

[38] R. J. Kociba, D. G. Keyes, J. E. Beyer, R. M. Carreon, C. E. Wade, D. A. Dittenber, and *et al.* Results of a 2-year chronic toxicity and oncogenicity study of 2,3,7,8-tetrachlorodibenzo-p-dioxin in rats. *Toxicology and Applied Pharmacology*, 46(2):279–303, 1978.

[39] L. Kuo and M. P. Cohen. Bayesian analysis for linearized multi-stage models in quantal bioassay. *Biometrical Journal*, 41(1):53–69, 1999.

[40] W. K. Lutz. Dose response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-non-linear curves, practical thresholds, j-shapes. *Mutation Research*, 405:117–124, 1998.

[41] H. Moon, H. Kim, J. J. Chen, and R. L. Kodell. Model averaging using the kullback information criterion in estimating effective doses for microbial infection and illness. *Risk Analysis*, 25:1147–1159, 2005.

[42] P. Mushak. Ad hoc and fast forward: The science of hormesis growth and development. *Environmental Health Perspectives*, 117:1333–1338, 2009.

[43] NTP. Carcinogenesis bioassay of 2,3,7,8-tetrachlorodibenzo-p-dioxin (cas no. 1746-01-6) in osborne-mendel rats and b6c3f1 mice (gavage study). Technical report, Department of Health and Human Services, Public Health Service, National Institutes of Health, National Toxicology Program, Research Triangle Park, NC, and Bethesda, MD, 1982.

[44] J. O'Quigley and M. Conaway. Continual reassessment and related dose-finding designs. *Statistical Science*, 25(2):202–216, 2010.

[45] J. O'Quigley, M. Pepe, and L. Fisher. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics*, 46, 1990.

[46] J. O'Quigley and L. Z. Shen. Continual reassessment method: A likelihood approach. *Biometrics*, 52:673–684, 1996.

[47] A. P. Oron and P. D. Hoff. Small-sample behavior of novel phase i cancer trial designs. *Clinical Trials*, 10:63–80, 2013.

[48] C. R. Palmer. Ethics and statistical methodology in clinical trials. *Journal of medical ethics*, 19:219–222, 1993.

[49] Z. Radak, H. Y. Chung, E. Koltai, A. W. Taylor, and S. Goto. Exercise, oxidative stress and hormesis. *Ageing Research Reviews*, 7(1):34–42, 2008.

[50] A. E. Raftery, D. Madigan, and J. A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92:179–191, 1997.

[51] H. Robbins and S. Monro. A stochatic approximation method. *Annals of Mathematical Statistics*, 29:351–356, 1951.

[52] F. J. Samaniego. *A comparison of the Bayesian and frequentist approaches to estimation*. New York: Springer, 2010.

[53] O. Schabenberger and J. B. Birch. Statistical dose-response models with hormetic effects. *Human and Ecological Risk Assessment: An International Journal*, 7(4):891–908, 2001.

[54] J. W. Seaman, J. D. Seaman, and J. D. Stamey. Hidden dangers of specifying noninformative priors. *The American Statistician*, 66(2):77–84, 2012.

[55] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[56] K. Shao and M. J. Small. Potential uncertainty reduction in model-averaged benchmark dose estimates informed by an additional dose study. *Risk Analysis*, 31:1156–1175, 2011.

[57] L. Z. Shen and J. O'Quigley. Consistency of continual reassessment method under model misspecification. *Biometrika*, 83:395–405, 1996.

[58] R. L. Sielken and D. E. Stevenson. Some implications for quantitative risk assessment if hormesis exists. *Human and Experimental Toxicology*, 17:259–262, 1998.

[59] R. Simon, B. Freidlin, L. Rubinstein, S. Arbuck, J. Collins, and M. C. Christian. Accelerated titration designs for phase i clinical trial in oncology. *Journal of the National Cancer Institute*, 89:1138–1147, 1997.

[60] B. E. Storer. Design and analysis of Phase I clinical trials. *Biometrics*, 45:925–937, 1989.

[61] K. A. Thayer, R. Melnick, K. Burns, D. Davis, and H. J. Fundamental flaws of hormesis for public health decisions. *Environmental Health Perspectives*, 113(10):1271–1276, 2005.

[62] M. Tighiouart, A. Rogatko, and J. S. Babb. Flexible Bayesian methods for cancer phase I clinical trials. Dose escalation with overdose control. *Statistics in Medicine*, 24(14):2183–2196, 2005.

[63] M. P. Waalkes, S. Rehm, C. W. Riggs, R. M. Bare, D. E. Devor, L. A. Poirier, M. L. Wenk, J. R. Henneman, and M. S. Balaschak. Cadmium carcinogenesis in male wistar [crl:(wi)br] rats: Dose-response analysis of tumor induction in the prostate and testes and at the injection site. *Cancer Research*, 48:4656–4663, 1998.

[64] G. B. Wetherill. Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society, Series B*, 25:1–48, 1963.

[65] M. W. Wheeler and A. J. Bailer. Properties of model-averaged bmdls: A study of model averaging in dichotomous response risk estimation. *Risk Analysis*, 27:659–670, 2007.

[66] G. B. Whetherill and H. Levitt. Sequential estimation of points of a psychometric function. *British Journal of Mathematical and Statistical Psychology*, 18:1–10, 1965.

[67] J. Whitehead and H. Brunier. Bayesian decision procedures for dose determining experiments (Disc: P897–899). *Statistics in Medicine*, 14:885–893, 1995.

[68] J. Whitehead and D. Williamson. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*, 8:445–467, 1998.

[69] J. Whitehead, Y. Zhou, J. Stevens, G. Blakey, and J. Price. Bayesian decision procedures for dose-escalation based on evidence of undesirable events and therapeutic benefit. *Statistics in Medicine*, 25:37–53, 2006.

[70] G. Yin and Y. Yuan. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association*, 104:954–968, 2009.

[71] G. A. Zapponi and I. Marcello. Low-dose risk, hormesis, analogical and logical thinking. *Annals New York Academy of Sciences*, 1076:839–857, 2006.

Appendix A

Appendices

A.1 Description of the 3 + 3 Design

A cohort of three patients is allocated to an arbitrarily low dose level. A trial continues until the following algorithm cannot find the next dose allocation among fixed experimental doses:

- If no AE is observed, escalate one dose level.
- If one AE is observed, stay at the same dose level.
- If two or more AEs are observed, de-escalate one dose level.
- At most six patients are allocated to a dose level.

The dose level with the observed proportion of AEs less than (or equal to) one-third is determined to be a MTD. A trial may be terminated without a determined MTD particularly when a small number of experimental doses is available.

A.2 Decomposition of the Loss Function in the IGM

Let $z_n^{(m)} = \sum_{i=1}^{n} \tau_i x_i^m$ for $m = 0, 1, 2$. The inverse of $\mathcal{I}_{n+1}$ is

$$\begin{aligned}\mathcal{I}_{n+1}^{-1} &= \begin{pmatrix} z_n^{(0)} + \tau_{n+1} & z_n^{(1)} + \tau_{n+1} x_{n+1} \\ z_n^{(1)} + \tau_{n+1} x_{n+1} & z_n^{(2)} + \tau_{n+1} x_{n+1}^2 \end{pmatrix}^{-1} \\ &= \Delta^{-1} \begin{pmatrix} z_n^{(2)} + \tau_{n+1} x_{n+1}^2 & -(z_n^{(1)} + \tau_{n+1} x_{n+1}) \\ -(z_n^{(1)} + \tau_{n+1} x_{n+1}) & z_n^{(0)} + \tau_{n+1} \end{pmatrix},\end{aligned}$$

with

$$\Delta = (z_n^{(0)} + \tau_{n+1})(z_n^{(2)} + \tau_{n+1} x_{n+1}^2) - (z_n^{(1)} + \tau_{n+1} x_{n+1})^2$$

and the loss function with respect to x_{n+1} for given $\vec{\beta}$ is

$$\begin{aligned}\mathcal{L}_I(x_{n+1} \mid \vec{\beta}) &= \nabla \vec{h}^T I_{n+1}^{-1} \nabla \vec{h} \\ &\propto \frac{(z_n^{(2)} + \tau_{n+1} x_{n+1}^2) - 2\,\mathrm{MTD}_\gamma (z_n^{(1)} + \tau_{n+1} x_{n+1}) + \mathrm{MTD}_\gamma^2\,(z_n^{(0)} + \tau_{n+1})}{(z_n^{(0)} + \tau_{n+1})\,(z_n^{(2)} + \tau_{n+1} x_{n+1}^2) - (z_n^{(1)} + \tau_{n+1} x_{n+1})^2}.\end{aligned}$$

By completing square, the numerator of $\mathcal{L}_I(x_{n+1} \mid \vec{\beta})$ is decomposed as

$$\begin{aligned}&\tau_{n+1}\,(x_{n+1} - \mathrm{MTD}_\gamma)^2 + z_n^{(0)} \left\{ \left(\mathrm{MTD}_\gamma - \frac{z_n^{(1)}}{z_n^{(0)}} \right)^2 + \frac{z_n^{(2)}}{z_n^{(0)}} - \left(\frac{z_n^{(1)}}{z_n^{(0)}} \right)^2 \right\} \\ &\qquad = \tau_{n+1}\,(x_{n+1} - \mathrm{MTD}_\gamma)^2 + z_n^{(0)} \left\{ (\mathrm{MTD}_\gamma - \mu_n)^2 + \sigma_n^2 \right\}.\end{aligned}$$

By expanding each term and completing square, the denominator of $\mathcal{L}_I(x_{n+1} \mid \vec{\beta})$ is decomposed as

$$\left\{ z_n^{(0)} z_n^{(2)} - (z_n^{(1)})^2 \right\} + z_n^{(0)} \tau_{n+1} \left\{ (x_{n+1} - \mu_n)^2 + \sigma_n^2 \right\}.$$

Putting the numerator and denominator together, the expression in Equation (3.2) is immediate.

A.3 Joint Prior Density Function

If we assume $\pi_0 \sim \text{Beta}(a_0, b_0)$, then the prior density of β_0 is

$$f(\beta_0) \propto (1 - e^{-\beta_0})^{a_0 - 1} e^{-b_0 \beta_0} \tag{A.1}$$

by the Jacobian transformation from π_0 to $\beta_0 = -\log(1 - \pi_0)$. We introduce a latent random variable $\zeta \mid \theta \sim \text{Bernoulli}(\theta)$ with $\theta \sim \text{Beta}(a_\theta, b_\theta)$ to indicate $\zeta = \mathbf{1}_{\beta_1 < 0}$, so that $\theta = P(\beta_1 < 0) = P(\Omega_H)$. The density functions of $\zeta \mid \theta$ and θ are

$$f(\zeta \mid \theta) = \theta^{\zeta}(1 - \theta)^{1 - \zeta} \tag{A.2}$$

and

$$f(\theta) \propto \theta^{a_\theta - 1}(1 - \theta)^{b_\theta - 1}\,, \tag{A.3}$$

respectively. We consider a prior density function for $\beta_1 \mid \zeta$ of the form

$$f(\beta_1 \mid \zeta) = \mathbf{1}_{\zeta = 1}\, f_1^-(\beta_1) + \mathbf{1}_{\zeta = 0}\, f_1^+(\beta_1)\,, \tag{A.4}$$

where $f_1^-(\cdot)$ and $f_1^+(\cdot)$ are some density functions with negative and nonnegative support, respectively. Conditional on β_0 and β_1, we define the prior density of β_2 as

$$f(\beta_2 \mid \beta_0, \beta_1) = \mathbf{1}_{\beta_1 < 0}\, f_2^{\Omega_H}(\beta_2) + \mathbf{1}_{\beta_1 \geq 0}\, f_2^+(\beta_2)\,, \tag{A.5}$$

where $f_2^+(\cdot)$ is some density function with nonnegative support and $f_2^{\Omega_H}(\cdot)$ is some density function which is defined on Ω_H given (β_0, β_1). One possible choice for $f_1^-(\cdot)$, $f_1^+(\cdot)$, $f_2^{\Omega_H}(\cdot)$, and $f_2^+(\cdot)$ is the class of truncated normal density functions, though any density functions with the appropriate support can be chosen.

Based upon the above likelihood function and prior density functions, the joint posterior density of $\vec{\beta}$, ζ, and θ is

$$f(\vec{\beta}, \zeta, \theta \mid \vec{y}) \propto L(\vec{\beta})\, f(\beta_0)\, f(\beta_1 \mid \zeta)\, f(\zeta \mid \theta)\, f(\theta)\, f(\beta_2 \mid \beta_1, \beta_0)\,.$$

Note that we have not introduced dependence between β_0 and β_1, but the two parameters are dependent in the posterior through $L(\vec{\beta})$. Gibbs sampling can be used to obtain realizations from the posterior distribution as presented in Appendix B.

A.4 Gibbs Sampling

Kuo and Cohen presented a Gibbs sampling algorithm for $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_M)$ in the multistage model assuming $\beta_m \geq 0$ for $m = 0, 1, \ldots, M$ [39]. However, the following algorithm is presented for direct application to our proposed model with the modified parameter space $\Omega = \Omega_M \cup \Omega_H$. Recall that the boundary of Ω_H depends on the values of the parameters. For the posterior sampling, we employ Gibbs sampling in the order of $\beta_0 \to \theta \to \zeta \to \beta_1 \to \beta_2$. We first initialize all parameters. One reasonable $\vec{\beta}^{(0)}$ is the OLSEs when we regress x and x^2 on

$$y^* = -\log(1 - \pi^*)\,, \;\; \pi^* = \frac{y + 0.5}{n + 1}\,.$$

For $t = 1, \ldots, T$, the Gibbs procedure proceeds as follows:

- *Step 1.* Sample $\beta_0^{(t)} \sim g(\beta_0)$ where

$$g(\beta_0) \propto L(\beta_0, \beta_1^{(t-1)}, \beta_2^{(t-1)})\, f(\beta_0)\, f(\beta_2^{(t-1)} \mid \beta_1^{(t-1)}, \beta_0)\,.$$

- *Step 2.* Sample $\theta^{(t)} \sim \text{Beta}(a_\theta + \zeta^{(t-1)}, b_\theta + 1 - \zeta^{(t-1)})$.

- *Step 3.* To sample $\zeta^{(t)} \in \{0, 1\}$ from a discrete distribution, first consider the following procedure.

  – Let $\zeta = 1$ and sample $(\beta_1, \beta_2) \sim g_1(\beta_1, \beta_2)$ where

$$g_1(\beta_1, \beta_2) \propto L(\beta_0^{(t)}, \beta_1, \beta_2)\, f_1^-(\beta_1)\, f_2^{\Omega_H}(\beta_2)\,.$$

  Call the samples $(\beta_1^{[1]}, \beta_2^{[1]})$ and evaluate $g_1(\beta_1^{[1]}, \beta_2^{[1]})$.

  – Let $\zeta = 0$ and sample $(\beta_1, \beta_2) \sim g_0(\beta_1, \beta_2)$ where

$$g_0(\beta_1, \beta_2) \propto L(\beta_0^{(t)}, \beta_1, \beta_2)\, f_1^+(\beta_1)\, f_2^+(\beta_2)\,.$$

  Call the samples $(\beta_1^{[0]}, \beta_2^{[0]})$ and evaluate $g_0(\beta_1^{[0]}, \beta_2^{[0]})$.

  Then sample $\zeta^{(t)} \in \{0, 1\}$ from the normalized probability mass function

$$P(\zeta = z) = \frac{g_z(\beta_1^{[z]}, \beta_2^{[z]})}{g_0(\beta_1^{[0]}, \beta_2^{[0]}) + g_1(\beta_1^{[1]}, \beta_2^{[1]})}$$

  for $z = 0, 1$.

- *Step 4.* No additional computational work is needed to sample $(\beta_1^{(t)}, \beta_2^{(t)})$. Let

$$\beta_j^{(t)} = \beta_j^{\left[\zeta^{(t)}\right]}$$

for $j = 1, 2$.

A.5 Connection to BMA

Let $P(\Omega_M)$ and $P(\Omega_H)$ denote assigned prior probabilities for each class of assumption with $P(\Omega_M) + P(\Omega_M) = 1$. By taking the product of Equations (A.1) to (A.5) and marginalizing over ζ and θ, it can be shown that the joint prior density of $\vec{\beta}$ can be written in the form of

$$f(\vec{\beta}) = P(\Omega_M)\, f_{\Omega_M}(\vec{\beta}) + P(\Omega_H)\, f_{\Omega_H}(\vec{\beta})\,,$$

where

$$P(\Omega_M) = \frac{\Gamma(a_\theta + b_\theta)}{\Gamma(a_\theta)\,\Gamma(b_\theta)}\,\frac{\Gamma(a_\theta)\,\Gamma(b_\theta + 1)}{\Gamma(a_\theta + b_\theta + 1)}$$

and

$$P(\Omega_H) = \frac{\Gamma(a_\theta + b_\theta)}{\Gamma(a_\theta)\,\Gamma(b_\theta)}\,\frac{\Gamma(a_\theta + 1)\,\Gamma(b_\theta)}{\Gamma(a_\theta + b_\theta + 1)}\,.$$

Then, the posterior density function of $\vec{\beta}$ under the two-stage model with the parameter space $\Omega = \Omega_M \cup \Omega_H$ is

$$\begin{aligned}
f(\vec{\beta} \mid \vec{y}) &= \frac{f(\vec{y} \mid \vec{\beta}) \left[P(\Omega_M)\, f_{\Omega_M}(\vec{\beta}) + P(\Omega_H)\, f_{\Omega_H}(\vec{\beta})\right]}{\int_\Omega f(\vec{y} \mid \vec{\beta}) \left[P(\Omega_M)\, f_{\Omega_M}(\vec{\beta}) + P(\Omega_H)\, f_{\Omega_H}(\vec{\beta})\right] d\vec{\beta}} \\
&= \frac{f(\vec{y} \mid \vec{\beta}) \left[P(\Omega_M)\, f_{\Omega_M}(\vec{\beta}) + P(\Omega_H)\, f_{\Omega_H}(\vec{\beta})\right]}{P(\Omega_M)\, f_{\Omega_M}(\vec{y}) + P(\Omega_H)\, f_{\Omega_H}(\vec{y})} \\
&= \frac{P(\Omega_M)\, f(\vec{y} \mid \vec{\beta})\, f_{\Omega_M}(\vec{\beta})}{P(\Omega_M)\, f_{\Omega_M}(\vec{y})} \frac{P(\Omega_M)\, f_{\Omega_M}(\vec{y})}{P(\Omega_M)\, f_{\Omega_M}(\vec{y}) + P(\Omega_H)\, f_{\Omega_H}(\vec{y})} \\
&\quad + \frac{P(\Omega_H)\, f(\vec{y} \mid \vec{\beta})\, f_{\Omega_H}(\vec{\beta})}{P(\Omega_H)\, f_{\Omega_H}(\vec{y})} \frac{P(\Omega_H)\, f_{\Omega_H}(\vec{y})}{P(\Omega_M)\, f_{\Omega_M}(\vec{y}) + P(\Omega_H)\, f_{\Omega_H}(\vec{y})} \\
&= f_{\Omega_M}(\vec{\beta} \mid \vec{y})\, P(\Omega_M \mid \vec{y}) + f_{\Omega_H}(\vec{\beta} \mid \vec{y})\, P(\Omega_H \mid \vec{y})\,,
\end{aligned}$$

where

$$P(\Omega_M \mid \vec{y}) = \frac{P(\Omega_M)\, f_{\Omega_M}(\vec{y})}{P(\Omega_M)\, f_{\Omega_M}(\vec{y}) + P(\Omega_H)\, f_{\Omega_H}(\vec{y})} = P(\beta_1 \geq 0 \mid \vec{y})$$

and

$$P(\Omega_H \mid \vec{y}) = \frac{P(\Omega_H)\, f_{\Omega_H}(\vec{y})}{P(\Omega_M)\, f_{\Omega_M}(\vec{y}) + P(\Omega_H)\, f_{\Omega_H}(\vec{y})} = P(\beta_1 < 0 \mid \vec{y})$$

are the posterior probability of monotonicity and hormesis, respectively. In other words, inference from the proposed two-stage model is equivalent to BMA. In the case of equal prior probabilities, $P(\Omega_M) = P(\Omega_H) = 0.5$, where $a_\theta = b_\theta$, the posterior probabilities simplify to $P(\Omega_M \mid \vec{y}) = f_{\Omega_M}(\vec{y}) \,/\, [f_{\Omega_M}(\vec{y}) + f_{\Omega_H}(\vec{y})]$ and $P(\Omega_H \mid \vec{y}) = f_{\Omega_H}(\vec{y}) \,/\, [f_{\Omega_M}(\vec{y}) + f_{\Omega_H}(\vec{y})]$, weighed by the marginal likelihood.

A.6 Detailed Expression of β_1-Loss

Let $\pi_i \equiv \pi_i(\beta)$ for concise notation. Let $w_i = (1-\pi_i)/\pi_i$ be an unnormalized weight and $z_i = x_i^{\beta_3} \in (0,1)$ be a dose in the transformed dose space by the exponetiation of β_3. Let $\Delta_i = \beta_1\beta_3(x_i^{\beta_3-1} + 2x_i^{2\beta_3-1})$. We further denote

$$s_m = \sum_{i=1}^{n} w_i z_i^m$$

for $m = 0, 1, 2, 3, 4$ and

$$t_{l,m} = \sum_{i=1}^{n} w_i \Delta_i^l z_i^m$$

for $l = 1, 2$ and $m = 0, 1, 2$. Then the expected Fisher information matrix is

$$\mathcal{I}(\vec{x}_n; \vec{\beta}) = \begin{pmatrix} s_0 & s_1 & s_2 & t_{10} \\ s_1 & s_2 & s_3 & t_{11} \\ s_2 & s_3 & s_4 & t_{12} \\ t_{10} & t_{11} & t_{12} & t_{02} \end{pmatrix}.$$

It can be shown that the numerator of the $(2,2)^{\text{th}}$ element of the inverted matrix is

$$\begin{aligned} & s_0(s_4 t_{20} - t_{12}^2) + s_2(t_{10}t_{12} - s_2 t_{20}) + t_{10}(s_2 t_{12} - s_4 t_{10}) \\ & \quad = \left(\sum_{i=1}^{n} w_i\right)\left\{\left(\sum_{i=1}^{n} w_i z_i^4\right)\left(\sum_{i=1}^{n} w_i \Delta_i^2\right) - \left(\sum_{i=1}^{n} w_i \Delta_i z_i^2\right)^2\right\} \\ & \quad + \left(\sum_{i=1}^{n} w_i z_i^2\right)\left\{\left(\sum_{i=1}^{n} w_i \Delta_i\right)\left(\sum_{i=1}^{n} w_i \Delta_i z_i^2\right) - \left(\sum_{i=1}^{n} w_i z_i^2\right)\left(\sum_{i=1}^{n} w_i \Delta_i^2\right)\right\} \\ & \quad + \left(\sum_{i=1}^{n} w_i \Delta_i\right)\left\{\left(\sum_{i=1}^{n} w_i z_i^2\right)\left(\sum_{i=1}^{n} w_i \Delta_i z_i^2\right) - \left(\sum_{i=1}^{n} w_i z_i^4\right)\left(\sum_{i=1}^{n} w_i \Delta_i\right)\right\}. \end{aligned}$$

The denominator is the determinant of $\mathcal{I}(\vec{x}_n; \vec{\beta})$.