

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Bayesian Model Averaging and Compromising in Dose-Response Studies

Steven B. Kim

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.68786>

Abstract

Dose-response models are applied to animal-based cancer risk assessments and human-based clinical trials usually with small samples. For sparse data, we rely on a parametric model for efficiency, but posterior inference can be sensitive to an assumed model. In addition, when we utilize prior information, multiple experts may have different prior knowledge about the parameter of interest. When we make sequential decisions to allocate experimental units in an experiment, an outcome may depend on decision rules, and each decision rule has its own perspective. In this chapter, we address the three practical issues in small-sample dose-response studies: (i) model-sensitivity, (ii) disagreement in prior knowledge and (iii) conflicting perspective in decision rules.

Keywords: dose-response models, model-sensitivity, model-averaging, prior-sensitivity, consensus prior, Bayesian decision theory, individual-level ethics, population-level ethics, Bayesian adaptive designs, sequential decisions, continual reassessment method, c-optimal design, Phase I clinical trials

1. Introduction

Dose-response modeling is often used to learn about the effect of an agent on a particular outcome with respect to dose. It is widely applied to animal-based cancer risk assessments and human-based clinical trials. A sample size is typically small; so many statistical issues can arise from a limited amount of data. The issues include the impact of a misspecified model, prior-sensitivity, and conflicting ethical perspectives in clinical trials. In this chapter, we focus on cases when an outcome variable of interest is binary (a predefined event happened or not) when an experimental unit is exposed to a dose. Main ideas are preserved for cases when an outcome variable is continuous or discrete.

There are two different approaches to statistical inference. One approach is called frequentist inference. In this framework, we often rely on the sampling distribution of a statistic and large-sample theories. Another approach is called Bayesian inference. It is founded on Bayes' Theorem, and it allows researchers to express prior knowledge independent of data. In a small-sample study, Bayesian inference can be more useful than frequentist inference because we can incorporate both researcher's prior knowledge and observed data to make inference for the parameter of interest. Bayesian ideas are briefly introduced for dose-response modeling with a binary outcome in Section 2.

In a small-sample study, we often rely on a parametric model to gain statistical efficiency (i.e., less variance in parameter estimation), but our inference can be severely biased by the use of a wrong model. To account for model uncertainty, it is reasonable to specify multiple models and make inference based on "averaged-inference." In this regard, Bayesian model averaging (BMA) is a useful method to gain robustness [1]. The BMA method has a wide range of application, and we focus its application to animal-based cancer risk assessments in Section 3.

In clinical trials, study participants are real patients, and therefore, we need to carefully consider ethics. There are conflicting perspectives of individual- and population-level ethics in early phase clinical trials. Individual-level ethics focuses on the benefit of trial participants, whereas population-level focuses on the benefit of future patients, which may require some level of sacrifice from trial participants. We compare the two conflicting perspectives in clinical trials based on Bayesian decision theory, and we discuss a compromising method in Section 4 [2, 3].

A sample size for an early phase (Phase I) clinical trial is often less than 30 subjects. Dose allocations for first few patients and statistical inference for future patients heavily depend on researcher's prior knowledge in sparse data. When multiple researchers have different prior knowledge about a parameter of interest, one compromising approach is to combine their prior elicitation and average them (i.e., consensus prior) [4, 5]. When we average the prior elicitation, there are two different approaches to determine the weight of each prior elicitation, weights determined before observing data and after observing data. We discuss operating characteristics of the two different weighting methods in the context of Phase I clinical trials in Section 5.

2. Bayesian inference

In statistics, we address a research question by a parameter, which is often denoted by θ . We begin Bayesian inference by modeling the prior knowledge about θ . A function, which models the prior knowledge about θ , is called the prior density function of θ , and we denote it by $f(\theta)$. It is a non-negative function, which satisfies $\int_{\Omega} f(\theta) d\theta = 1$, where Ω is the set of all possible values of θ (i.e., parameter space). We then model data $\vec{y} = (y_1, \dots, y_n)$ given θ . The likelihood function, denoted by $f(\vec{y}|\theta)$, quantifies the likelihood of observing a particular

sample $\vec{y} = (y_1, \dots, y_n)$ under an assumed probability model. By Bayes' Theorem, we update our knowledge about θ after observing data \vec{y} as

$$f(\theta|\vec{y}) = \frac{f(\vec{y}|\theta)f(\theta)}{f(\vec{y})}. \quad (1)$$

The function $f(\theta|\vec{y})$ is called the posterior density function of θ given data \vec{y} . Since we treat observed data $\vec{y} = (y_1, \dots, y_n)$ as fixed numbers, we often express Eq. (1) as follows

$$f(\theta|\vec{y}) \propto f(\vec{y}|\theta) f(\theta) = k f(\vec{y}|\theta) f(\theta), \quad (2)$$

where k is the normalizing constant which makes $\int_{\Omega} f(\theta|\vec{y}) d\theta = 1$. We can often realize $f(\theta|\vec{y})$ based on the prior density function $f(\theta)$ and the likelihood function $f(\vec{y}|\theta)$ without considering the denominator $f(\vec{y}) = \int f(y|\theta)f(\theta) d\theta$ in Eq. (1) which is called the marginal likelihood.

2.1. Example

Suppose we observe $n = 20$ rats for 2 years. Let π be the parameter of interest, which is interpreted as the probability of developing some type of tumor. Suppose a researcher models the prior knowledge about π using the prior density function

$$f(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}, \quad 0 < \pi < 1. \quad (3)$$

It is known as the beta distribution with shape parameters $a > 0$ and $b > 0$. We often denote the beta distribution by $\pi \sim \text{Beta}(a, b)$, and the values of a and b must be specified by the researcher independent from data. Let $\vec{y} = (y_1, \dots, y_n)$ denote observed data, where $y_i = 1$ if the i^{th} rat developed tumor and $y_i = 0$ otherwise. Assuming y_1, \dots, y_n are independent observations, the likelihood function is as follows

$$f(\vec{y}|\pi) = \prod_{i=1}^n \pi^{y_i} (1-\pi)^{1-y_i} = \pi^s (1-\pi)^{n-s}, \quad (4)$$

where $s = \sum_{i=1}^n y_i$ is the total number of rats developed tumor. By Eq. (2), the posterior density function of π is as follows

$$f(\pi|\vec{y}) = k \pi^{a+s-1} (1-\pi)^{b+n-s-1}, \quad (5)$$

where $k = \frac{\Gamma(a+b+n)}{\Gamma(a+s)\Gamma(b+n-s)}$ is the normalizing constant, which makes $\int_0^1 f(\pi|\vec{y}) d\pi = 1$. We can recognize that $\pi|\vec{y} \sim \text{Beta}(a+s, b+n-s)$.

If the researcher fixed $a = 2$ and $b = 3$ and observed $s = 9$ from a sample of size $n = 20$, the prior density function is $f(\pi) = k\pi(1 - \pi)^2$ with $k = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} = 12$, and the posterior density function is $f(\pi|\vec{y}) = k\pi^{10}(1 - \pi)^{13}$ with $k = \frac{\Gamma(25)}{\Gamma(11)\Gamma(14)} = 27457584$. The prior and posterior distributions are shown in **Figure 1**. The knowledge about π becomes more certain (less variance) after observing the data.

2.2. Example

This example is simplified from Shao and Small [6]. In dose-response studies, we model π as a function of dose x . There are many link functions between π and x used in practice. In this example, we focus on a link function

$$\pi_x = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (6)$$

which is known as a logistic regression model. It is commonly assumed that a dose-response curve increases with respect to dose, so we assume $\beta_1 > 0$ (and β_0 can be any real number). There are two regression parameters in Eq. (6), β_0 and β_1 , and we denote them as $\vec{\beta} = (\beta_0, \beta_1)$. **Figure 2** presents two dose-response curves. The solid curve is generated by $\vec{\beta} = (-1, 2)$, and the dotted curve is generated by $\vec{\beta} = (-2, 5)$. As β_0 increases, the background risk $\pi_0 = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$ increases, where π_0 is interpreted as the probability of tumor development at dose $x = 0$. The dose-response curve increases when $\beta_1 > 0$, and it decreases when $\beta_1 < 0$. The rate of change in the dose-response curve is determined by $|\beta_1|$.

To express prior knowledge about $\vec{\beta}$, we need to find an appropriate prior density function $f(\vec{\beta})$. It is not simple because it is difficult to express one's knowledge on the two-dimensional parameters $\vec{\beta} = (\beta_0, \beta_1)$. For mathematical convenience, some practitioners use a flat prior density function $f(\vec{\beta}) \propto 1$. Another way of expressing a lack of prior knowledge about $\vec{\beta}$ is as follows

$$f(\vec{\beta}) \propto \frac{1}{2\pi\sigma^2} e^{-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}} \mathbf{I}_{\beta_1 > 0} \quad (7)$$

with an arbitrarily large value of σ [6]. When a reliable source of prior information is available, there is a practical method, which is known as the conditional mean prior [7], and it will be discussed in a later section (see Section 4.2). In an experiment, the experimental doses $\vec{x} = (x_1, \dots, x_n)$ are fixed, and we observe random binary outcomes $\vec{y} = (y_1, \dots, y_n)$. Given \vec{y} (and fixed \vec{x}), the likelihood function is as follows

$$f(\vec{y}|\vec{\beta}) = \prod_{i=1}^n \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{1 - y_i} = \frac{e^{\beta_0 s_1 + \beta_1 s_2}}{\prod_{i=1}^n (1 + e^{\beta_0 + \beta_1 x_i})}, \quad (8)$$

Prior and Posterior Distributions

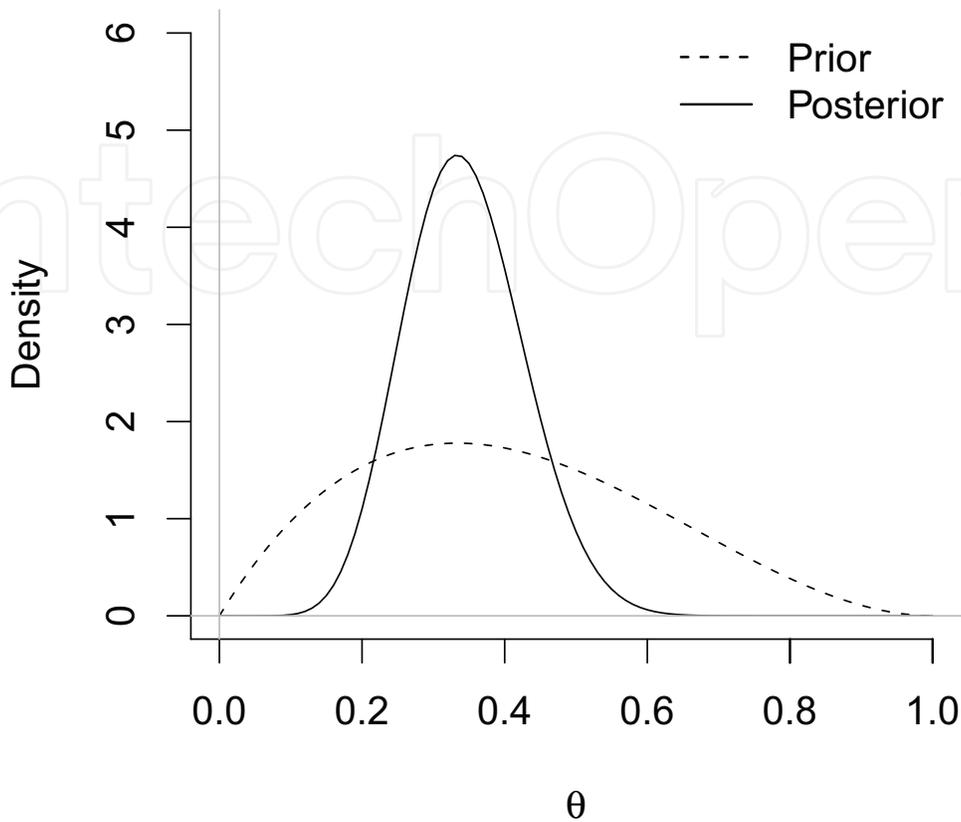


Figure 1. The prior $f(\pi)$ in the dotted curve and the posterior $f(\pi|\vec{y})$ in the solid curve.

where $s_1 = \sum_{i=1}^n y_i$ and $s_2 = \sum_{i=1}^n x_i y_i$. By incorporating both prior and data, the posterior density function is as follows

$$f(\vec{\beta} | \vec{y}) \propto f(\vec{\beta}) \frac{e^{\beta_0 s_1 + \beta_1 s_2}}{\prod_{i=1}^n (1 + e^{\beta_0 + \beta_1 x_i})} \tag{9}$$

In an animal-based studies, one parameter of interest is the median effective dose, which is denoted by ED_{50} . It is the dose, which satisfies

$$\pi_{ED_{50}} = \frac{e^{\beta_0 + \beta_1 ED_{50}}}{1 + e^{\beta_0 + \beta_1 ED_{50}}} = .5, \tag{10}$$

and it can be shown that $ED_{50} = -\frac{\beta_0}{\beta_1}$ by algebra. In the case of $\beta_0 = -2$ and $\beta_1 = 5$, we have $ED_{50} = .4$ as describe in the figure with the dotted curve. In the case of $\beta_0 = -1$ and $\beta_2 = 2$, we have $ED_{50} = .5$ as described in the figure with the solid curve.

In 1997, International Agency for Research on Cancer classified 2,3,7,8-Tetrachlorodibenzo-p-dioxin (known as TCDD) as a carcinogen for humans based on various empirical evidence [8].

Dose-Response Curves (Logistic Link)

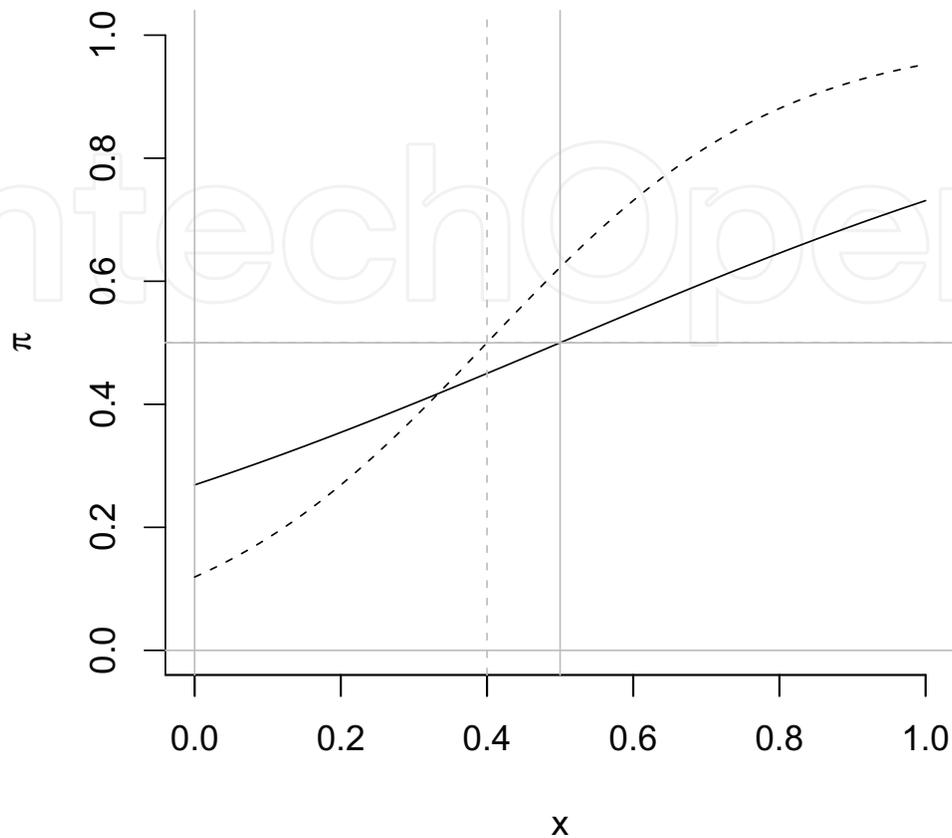


Figure 2. Two dose-response curves using the logistic link.

In 1978, Kociba et al. presented the data on male Sprague-Dawley rats at four experimental doses 0, 1, 10 and 100 nanograms per kilogram per day (ng/kg/day) [9]. In the control dose group, nine of 86 rats developed tumor (known as hepatocellular carcinoma); three of 50 rats developed the tumor at dose 1; 18 of 50 rats developed the tumor at dose 10; and 34 of 48 rats developed the tumor at dose 100 [6]. Without loss of generality, we let $x_i = 0$ for $i = 1, \dots, 86$; $x_i = 1$ for $i = 87, \dots, 136$; $x_i = 10$ for $i = 137, \dots, 186$; and $x_i = 100$ for $i = 187, \dots, 234$. The given information is sufficient to calculate $s_1 = \sum_{i=1}^n y_i = 64$ and $s_2 = \sum_{i=1}^n x_i y_i = 3583$. By the use of the flat prior $f(\vec{\beta}) \propto 1$ with the restriction $\beta_1 > 0$, given the observed sample of size $n = 234$, we can generate random numbers of $\vec{\beta} = (\beta_0, \beta_1)$ from the posterior density function

$$f(\vec{\beta} | \vec{y}) \propto \frac{e^{\beta_0 s_1 + \beta_1 s_2}}{\prod_{i=1}^n (1 + e^{\beta_0 + \beta_1 x_i})} I_{\beta_1 > 0}, \quad (11)$$

where $I_{\beta_1 > 0} = 1$ if $\beta_1 > 0$ and $I_{\beta_1 > 0} = 0$ otherwise. Using a method of Markov Chain Monte Carlo (MCMC), we can approximate the posterior distribution of $\vec{\beta}$ as shown in the left panel of **Figure 3**. By transforming (β_0, β_1) to $ED_{50} = -\frac{\beta_0}{\beta_1}$, we can approximate the posterior

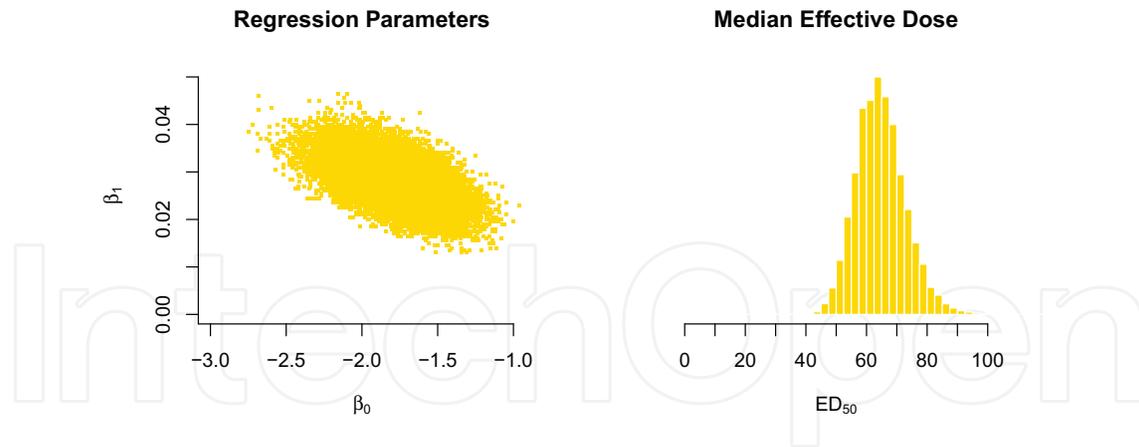


Figure 3. Approximate posterior distributions of (β_0, β_1) and $ED_{50} = -\frac{\beta_0}{\beta_1}$.

distribution of the median effective dose ED_{50} as shown in the right panel. The posterior mean of ED_{50} is $E(ED_{50}|\vec{y}) = 64.9$ with 95% credible interval (50.8, 82.5), the 2.5th percentile and the 97.5th percentile of the posterior distribution.

3. Bayesian model averaging

In a small sample, we borrow the strength of a parametric model to gain efficiency in parameter estimation. However, an assumed model may not describe the true dose-response relationship adequately. The impact of model misspecification is not negligible particularly in a poor experimental design. In such a limited practical situation, Bayesian model averaging (BMA) can be a useful method to account for model uncertainty. It is widely applied in practice, and in this section, we focus on the application to cancer risk assessment for the estimation of a benchmark dose [1, 6, 10, 11].

Let θ denote a parameter of interest. Suppose we have a set of K candidate models denoted by $\mathcal{M} = \{M_1, \dots, M_K\}$. Let $\vec{\beta}_k$ denote the vector of regression parameters under model M_k for $k=1, \dots, K$. Suppose θ is a function of $\vec{\beta}_k$, and the interpretation of θ must be common across all models. Let $f(\vec{\beta}_k|M_k)$ and $f(\vec{y}|\vec{\beta}_k, M_k)$ denote the prior density function and the likelihood function, respectively, under M_k . By the Law of Total Probability, the posterior density function of θ is as follows

$$f(\theta|\vec{y}) = \sum_{k=1}^K f(\theta|M_k, \vec{y}) P(M_k|\vec{y}). \quad (12)$$

In Eq. (12), the posterior density function $f(\theta|M_k, \vec{y})$ depends on model M_k , and the posterior model probability $P(M_k|\vec{y})$ quantifies the plausibility of model M_k after observing data, which is given by

$$P(M_k|\vec{y}) = \frac{f(\vec{y}|M_k) P(M_k)}{\sum_{j=1}^K f(\vec{y}|M_j) P(M_j)}. \quad (13)$$

In Eq. (13), the prior model probability $P(M_k)$ is determined before observing data such that $P(M_k) > 0$ for $k = 1, \dots, K$ and $\sum_{k=1}^K P(M_k) = 1$. The marginal likelihood under M_k requires the integration

$$f(\vec{y}|M_k) = \int f(\vec{y}|\vec{\beta}_k, M_k) f(\vec{\beta}_k|M_k) d\vec{\beta}_k. \quad (14)$$

In the BMA method, all K models contribute to inference of θ through the averaged posterior density function in Eq. (12), and the weight of contribution is determined by Bayes' Theorem in Eq. (13).

3.1. Example

This example is continued from the example in Section 2.2. Recall π_x is interpreted as the probability of a toxic event (tumor development) at dose x . In many cancer risk assessments, a parameter of interest is θ_γ at a fixed risk level γ , which is defined as follows

$$\gamma = \frac{\pi_{\theta_\gamma} - \pi_0}{1 - \pi_0} \quad (15)$$

or equivalently $\pi_{\theta_\gamma} = \pi_0 + (1 - \pi_0)\gamma$. In words, θ_γ is a dose corresponding to a fixed increase in the risk level. In frequentist framework, Crump defined a benchmark dose as a lower confidence limit for θ_γ [12]. In Bayesian framework, an analogous definition would be a lower credible bound (i.e., a fixed low percentile of the posterior distribution of θ_γ). The definition is widely applied to the public health protection [13].

In practice, γ is fixed between 0.01 and 0.1. Often, the estimation of θ_γ is highly sensitive to an assumed dose-response model because we have a lack of information at low doses. Shao and Small fixed $\gamma = 0.1$ and applied BMA with $K = 2$ models, logistic model and quantal-linear model [6]. In the quantal-linear model, the probability of tumor development is modeled by

$$\pi_x = \beta_0 + (1 - \beta_0)(1 - e^{-\beta_1 x}). \quad (16)$$

with the restrictions $0 < \beta_0 < 1$ and $\beta_1 > 0$ under the monotonic assumption. The logistic model was given in Eq. (6) of Section 2.2.

Let M_1 denote the logistic model, and let M_2 denote the quantal-linear model. Assume the uniform prior model probabilities $P(M_1) = P(M_2) = .5$ and flat priors on the regression parameters. By posterior sampling, we can approximate the posterior model probabilities $P(M_1|\vec{y}) = .049$ and $P(M_2|\vec{y}) = .951$. Under M_1 , the posterior mean of $\theta_{0.1}$ is 20.95 with the 5th percentile 16.74. Under M_2 , the posterior mean is 8.25 with the 5th percentile 5.95. These

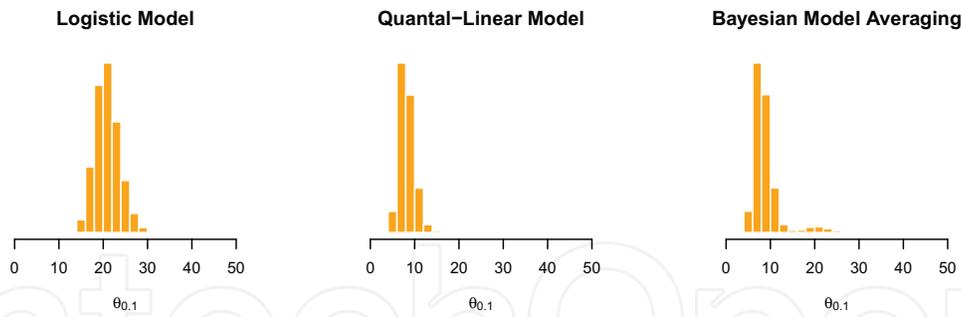


Figure 4. Posterior distributions of $\theta_{0.1}$ from the logistic model (left panel), the quantal-linear model (middle panel), and the Bayesian model averaging (right panel).

results are very similar to the results reported by Shao and Small [6]. From these model-specific statistics, we can calculate the model-averaged posterior mean

$$E(\theta_{0.1}|\vec{y}) = \sum_{k=1}^2 E(\theta_{0.1}|M_k, \vec{y}) P(M_k|\vec{y}) = 20.95 (.049) + 8.25 (.951) = 8.87. \quad (17)$$

However, we are not able to calculate the 5th percentile of the model-averaged posterior distribution based on the given statistics. In fact, we need to approximate the posterior distribution $f(\theta_{0.1}|\vec{y})$, which is a mixture of $f(\theta_{0.1}|M_1, \vec{y})$ and $f(\theta_{0.1}|M_2, \vec{y})$ weighted by $P(M_1|\vec{y}) = .049$ and $P(M_2|\vec{y}) = .951$, respectively, as shown in **Figure 4**. In the figure, the left panel shows an approximation of $f(\theta_{0.1}|M_1, \vec{y})$, the middle panel shows an approximation of $f(\theta_{0.1}|M_2, \vec{y})$, and the right panel shows an approximation of the averaged posterior $f(\theta_{0.1}|\vec{y})$. The averaged posterior density $f(\theta_{0.1}|\vec{y})$ is bimodal, but it is very close to $f(\theta_{0.1}|M_2, \vec{y})$ because the quantal-linear model M_2 fits the data better than the logistic model M_1 by a Bayes factor of $\frac{P(M_2|\vec{y})}{P(M_1|\vec{y})} = \frac{.951}{.049} = 19.4$. The 5th percentile of the model-averaged posterior distribution is approximately 5.97, and it is a BMA-BMD based on the BMA method proposed by Raftery et al. [1] and the BMD estimation method suggested by Crump [12].

4. Application of Bayesian decision theory to Phase I trials

In a Phase I cancer trial, the main objectives are to study the safety of a new chemotherapy and to determine an appropriate dose for future patients. Since trial participants are cancer patients, dose allocations require ethical considerations. Whitehead and Williams discussed several Bayesian approaches to dose allocations [14]. One decision rule is devised from the perspective of trial participants (individual-level ethics), and another decision rule is devised from the perspective of future patients (population-level ethics). However, a decision rule, which is devised from the population-level ethics, is not widely accepted in current practice [15]. Instead, there are some proposed decision rules, which compromise between the individual- and population-level perspectives [3, 16]. In this section, we discuss the two

conflicting perspectives in Phase I clinical trials and a compromising method based on Bayesian decision theory.

Assume a dose-response relationship follows a logistic model

$$\pi_x = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \quad (18)$$

where x is a dose in the logarithmic scale (base e) and π_x is the probability of observing an adverse event due to the toxicity of a new chemotherapy at dose x . The logarithmic transformation on the dose is to satisfy $\pi_x \rightarrow 0$ as $x \rightarrow 0$. Let $\vec{x}_n = (x_1, \dots, x_n)$ denote a series of decisions for n patients (i.e., allocated doses) and $\vec{y}_n = (y_1, \dots, y_n)$ denote a series of observed responses, where $y_i = 1$ indicates an adverse event and $y_i = 0$ otherwise. Let $L(\vec{\beta}, x_{n+1})$ denote a loss by allocating the next patient at x_{n+1} . Based on Bayesian Decision Theory, we want to find x_{n+1} which minimizes the posterior mean of $L(\vec{\beta}, x_{n+1})$. If we let \mathcal{A} denote an action space, a set of all possible dose allocations for the next patients, the decision rule can be written as follows:

$$x_{n+1}^* = \operatorname{argmin}_{x_{n+1} \in \mathcal{A}} E\left(L(\vec{\beta}, x_{n+1}) \mid \vec{y}_n\right). \quad (19)$$

A choice of L has a substantial impact on the operating characteristics of a Phase I trial including (i) the degree of under- and over-dosing in trial, (ii) the observed number of adverse events at the end of a trial, and (iii) the quality of estimation at the end of a trial.

4.1. Parameter of interest: maximum tolerable dose

Let N denote an available sample size for a Phase I clinical trial. A typical sample size is $N \leq 30$. Let γ denote a target risk level, the probability of an adverse event. In a cancer study, a typical target risk level γ is fixed between .15 and .35 depending on the severity of an adverse event. Then, the dose corresponding to γ is called a maximum tolerable dose (MTD) at level γ , and we denote it by θ_γ in the logarithmic scale. Under the logistic model in Eq. (18), it is defined as follows

$$\theta_\gamma = \frac{\log\left(\frac{\gamma}{1-\gamma}\right) - \beta_0}{\beta_1}. \quad (20)$$

At the end of a trial (observing N responses), we estimate θ_γ by the posterior mean $\hat{\theta}_{\gamma, N} = E(\theta_\gamma \mid \vec{y}_N)$ for future patients.

4.2. Prior density function: conditional mean priors

A consequence of sequential decisions heavily depends on a prior density function $f(\vec{\beta})$. In particular, the first decision x_1 must be made based on prior knowledge only because empirical evidence is not observed yet. In addition, the later decisions x_2, x_3, \dots and the final inference of

θ_γ are substantially affected by $f(\vec{\beta})$ as a Phase I study is typically based on a small sample. In this regard, we want to carefully utilize researchers' prior knowledge about $\vec{\beta}$, but it may be difficult to express their prior knowledge directly through $f(\vec{\beta})$. In this section, we discuss a method of eliciting prior knowledge, which is more tractable than prior elicitation directly on $\vec{\beta}$.

Suppose a researcher selects two arbitrarily doses, say $x_{-1} < x_0$. Then, the researcher may express their prior knowledge by two independent beta distributions

$$\pi_{x_i} = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \sim \text{Beta}(a_i, b_i), \quad j = -1, 0. \quad (21)$$

Using the Jacobian transformation from $(\pi_{x_{-1}}, \pi_{x_0})$ to $\vec{\beta} = (\beta_0, \beta_1)$, it can be shown that the prior density function of $\vec{\beta}$ is given by

$$f(\vec{\beta}) \propto (x_0 - x_{-1}) \prod_{i=-1}^0 \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{a_i} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{b_i}. \quad (22)$$

It is known as conditional mean priors under the logistic model [7].

4.3. Posterior density function: conjugacy

For notational convenience, we let $y_i = a_i$ and $n_i = a_i + b_i$ for $i = -1, 0$. By conjugacy, the posterior density function of $\vec{\beta}$ can be concisely written as follows

$$f(\vec{\beta} | \vec{y}_n) \propto \frac{e^{\beta_0 s_1 + \beta_1 s_2}}{\prod_{i=-1}^0 (1 + e^{\beta_0 + \beta_1 x_i})^{n_i}}, \quad (23)$$

where $s_1 = \sum_{i=-1}^0 y_i$ and $s_2 = \sum_{i=-1}^0 x_i y_i$. After observing n responses, the decision rule for the next patient is as follows

$$x_{n+1}^* = \operatorname{argmin}_{x_{n+1} \in \mathcal{A}} \int L(\vec{\beta}, x_{n+1}) f(\vec{\beta} | \vec{y}_n) d\vec{\beta}. \quad (24)$$

4.4. Loss functions for individual- and population-level ethics

A loss function, which reflects the perspective of individual-level ethics, is as follows:

$$L_I(\vec{\beta}, x_{n+1}) = (x_{n+1} - \theta_\gamma)^2. \quad (25)$$

This loss function is analogous to the original continual reassessment method proposed by O'Quigley et al. [17]. The square error loss attempts to treat a trial participant at θ_γ , and the expected square error loss is minimized by the posterior mean of θ_γ .

From the perspective of population-level ethics, Whitehead and Brunier proposed a loss function, which is equal to the asymptotic variance of the maximum likelihood estimator for θ_γ [18]. The Fisher expected information matrix with a sample of size $n + 1$ is given by

$$\mathcal{I}(\vec{\beta}) = \begin{pmatrix} \sum_{i=1}^{n+1} \tau_i & \sum_{i=1}^{n+1} \tau_i x_i \\ \sum_{i=1}^{n+1} \tau_i x_i & \sum_{i=1}^{n+1} \tau_i x_i^2 \end{pmatrix}, \quad (26)$$

where $\tau_i = \pi_{x_i}(1 - \pi_{x_i})$. Then, the loss function (the asymptotic variance) is given by

$$L_P(\vec{\beta}, x_{n+1}) = [\nabla \vec{h}(\vec{\beta})]^T [\mathcal{I}(\vec{\beta})]^{-1} [\nabla \vec{h}(\vec{\beta})], \quad (27)$$

where

$$\nabla \vec{h}(\vec{\beta}) = \begin{pmatrix} \frac{\partial \theta_\gamma}{\partial \beta_0} \\ \frac{\partial \theta_\gamma}{\partial \beta_1} \end{pmatrix} = -\frac{1}{\beta_1} \begin{pmatrix} 1 \\ \theta_\gamma \end{pmatrix} \quad (28)$$

is the gradient vector, the partial derivatives of θ_γ with respect to β_0 and β_1 . Kim and Gillen decomposed the population-level loss function as follows

$$L_P(\vec{\beta}, x_{n+1}) = \frac{\tau_{n+1}(x_{n+1} - \theta_\gamma)^2 + s_n^{(0)} [(\theta_\gamma - \mu_n)^2 + \sigma_n^2]}{\left[s_n^{(0)} s_n^{(2)} - s_n^{(1)} s_n^{(1)} \right] + s_n^{(0)} \tau_{n+1} [(x_{n+1} - \mu_n)^2 + \sigma_n^2]}, \quad (29)$$

where

$$\begin{aligned} s_n^{(m)} &= \sum_{i=1}^n \tau_i x_i^m, \quad m = 0, 1, 2, \\ \mu_n &= \sum_{i=1}^n w_i x_i, \\ \sigma_n^2 &= \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i \right)^2 \end{aligned} \quad (30)$$

with the weight defined as $w_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i}$ [3]. Eq. (29) has the following important remarks. In

fact, $L_P(\vec{\beta}, x_{n+1})$ considers individual-level ethics by including $L_I(\vec{\beta}, x_{n+1}) = (x_{n+1} - \theta_\gamma)^2$ in the numerator. By including $(x_{n+1} - \mu_n)^2$ in the denominator, where $\mu_n = \sum_{i=1}^n w_i x_i$, the population-level loss function reduces a loss by allocating the next patient further away from the weighted average of previously allocated doses (i.e., devised from information gain). In long run, $L_P(\vec{\beta}, x_{n+1})$ is devised from a compromise between individual- and population-level

ethics, but the compromising process is rather too slow to be implemented in a small-sample Phase I clinical trial [3].

4.5. Loss function for compromising the two perspectives

Kim and Gillen proposed to accelerate the compromising process by modifying $L_P(\vec{\beta}, x_{n+1})$ of Eq. (29) as follows

$$L_{B,\lambda}(\vec{\beta}, x_{n+1}) = \frac{a_n(\lambda) \tau_{n+1} (x_{n+1} - \theta_\gamma)^2 + s_n^{(0)} [(\theta_\gamma - \mu_n)^2 + \sigma_n^2]}{[s_n^{(0)} s_n^{(2)} - s_n^{(1)}] + s_n^{(0)} \tau_{n+1} [(x_{n+1} - \mu_n)^2 + \sigma_n^2]}, \quad (31)$$

where

$$a_n(\lambda) = \left(1 + \frac{n}{N}\right)^\lambda \left(1 + \frac{\sum_{i=1}^n y_i}{N_\gamma}\right) \quad (32)$$

is an accelerating factor [3]. It has two implications. First, the compromising process is accelerated toward the individual-level ethics as the trial proceeds (i.e., n increases). Second, the compromising process toward the individual-level ethics is accelerated at a faster rate when an adverse event is observed (i.e., $\sum_{i=1}^n y_i$ increases). The tuning parameter λ controls the rate of acceleration. It imposes more emphasis on population-level ethics as $\lambda \rightarrow 0$ and more emphasis on individual-level ethics as $\lambda \rightarrow \infty$. The choice of λ shall depend on the severity level of an adverse event.

4.6. Simulation

To study the operating characteristics of $L_{B,\lambda}$ with respect to λ , we assume the logistic model with $\beta_0 = -3$ and $\beta_1 = .8$ as a true dose-response relationship as shown in **Figure 5** in the left panel. The target risk level is fixed at $\gamma = .2$, so the true MTD is given by $\theta_{.2} = 2.02$ in the logarithmic scale. We consider three different priors based on the conditional mean priors given in Eq. (22). For simplicity, we set $a_{-1} = 1$, $b_{-1} = 3$, $a_0 = 3$ and $b_0 = 1$ for all three priors. Then, we let $x_{-1} = -4$ and $x_0 = 4$ for Prior 1; $x_{-1} = 0$ and $x_0 = 8$ for Prior 2; and $x_{-1} = 4$ and $x_0 = 12$ for Prior 3. **Figure 5** in the right panel shows an approximated $f(\theta_{.2})$ for each prior. Prior 1 significantly underestimates the true $\theta_{.2} = 2.02$ with prior mean $E(\theta_{.2}) = -1.70$, Prior 3 overestimates the truth with $E(\theta_{.2}) = 5.38$, and Prior 2 has a prior estimate relatively close to the truth with $E(\theta_{.2}) = 1.40$.

Let $N = 20$ be a fixed sample size. Let $Y_i = 1$ denote an adverse event observed from the i th patient ($Y_i = 0$ otherwise), so $\sum_{i=1}^N Y_i$ denotes the total number of adverse events observed at the end of a trial. The sum $\sum_{i=1}^n Y_i$ is random from a trial to another trial, and we want $\sum_{i=1}^n Y_i$ to behave like Binomial(20, .2) which is the case when we treat $N = 20$ to the true MTD $\theta_{.2}$. **Figure 6** shows three simulated trials under the loss function $L_{B,\lambda}$ with $\lambda = 0, 1, 5$. When $\lambda = 0$,

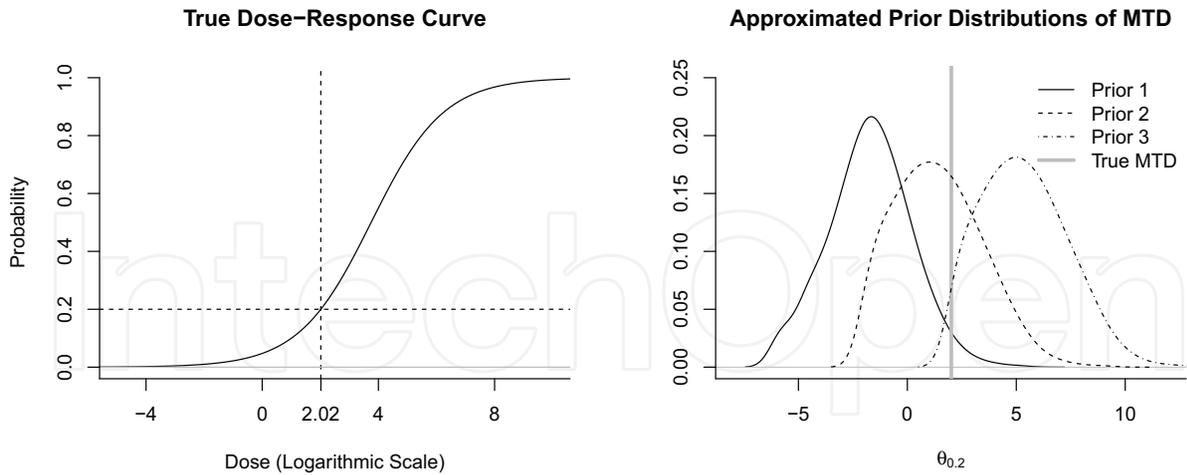


Figure 5. The true dose-response relationship $\pi_x = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ with $\beta_0 = -3$ and $\beta_1 = .8$ (where x is the dose in the logarithmic scale) in the simulation (left panel) and the three prior distributions of $\theta_{.2}$ approximated by kernel density (right panel).

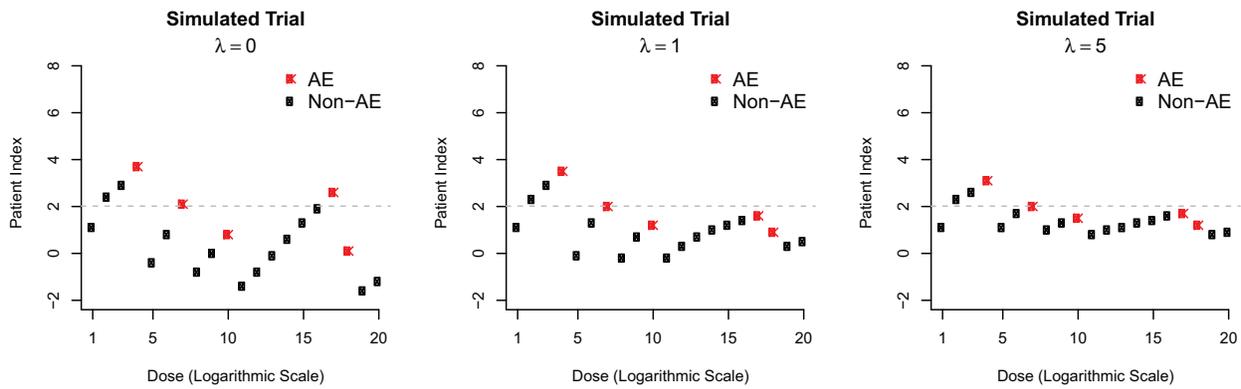


Figure 6. Three simulated trials using the loss function $L_{B,\lambda}$ with $\lambda = 0$ (left), $\lambda = 1$ (middle) and $\lambda = 5$ (right) with a sample of size $N = 20$ and assumed parameter values $\beta_0 = -3$, $\beta_1 = 8$ and $\theta_{.2} = 2.02$.

the up-and-down scheme has a high degree of fluctuation in order to maximize information about $\theta_{.2}$. When $\lambda = 1$, the up-and-down scheme is stabilized after the first few adverse events, and the stabilization occurs quickly when $\lambda = 5$ to treat trial participants near an estimated $\theta_{.2}$.

Let $\hat{\theta}_{.2} = E(\theta_{.2} | \vec{y}_N)$, the posterior estimate of $\theta_{.2}$ at the end of a trial, so $\pi_{\hat{\theta}_{.2}}$ implies the true probability of an adverse event at the estimated MTD. We focus on the following criteria: (i) $E(\pi_{\hat{\theta}_{.2}})$ which we desire to be close to $\gamma = .2$ for future patients, (ii) $V(\pi_{\hat{\theta}_{.2}})$ which we desire to be as low as possible for future patients, (iii) $E[(\pi_{\hat{\theta}_{.2}} - .2)^2]$ which we desire to be as low as possible for future patients, (iv) $E(\sum_{i=1}^{20} Y_i)$ which we desire to be close to $N\gamma = 4$ for trial participants and (v) $P(3 \leq \sum_{i=1}^{20} Y_i \leq 5)$ which we desire to be close to one for trial participants.

Prior	λ	$E(\pi_{\hat{\theta},2})$	$V(\pi_{\hat{\theta},2})$	$E[(\pi_{\hat{\theta},2} - .2)^2]$	$E(\sum_{i=1}^{20} Y_i)$	$P(3 \leq \sum_{i=1}^{20} Y_i \leq 5)$
1	0	0.0964	0.0019	0.0126	2.4353	0.4318
	.5	0.1034	0.0024	0.0118	2.0997	0.2298
	1	0.1082	0.0028	0.0113	1.8969	0.1714
	2	0.1100	0.0031	0.0112	1.6929	0.1211
	5	0.1157	0.0035	0.0106	1.3128	0.0596
2	0	0.1665	0.0054	0.0065	4.1217	0.9889
	.5	0.1705	0.0056	0.0065	3.9598	0.9877
	1	0.1727	0.0060	0.0068	3.9025	0.9670
	2	0.1751	0.0066	0.0072	3.8707	0.9291
	5	0.1763	0.0067	0.0073	3.8442	0.9068
3	0	0.2743	0.0048	0.0103	6.1875	0.1600
	.5	0.2673	0.0048	0.0093	6.3954	0.1430
	1	0.2606	0.0046	0.0083	6.6194	0.1165
	2	0.2562	0.0045	0.0077	6.8035	0.1020
	5	0.2499	0.0044	0.0068	7.0274	0.0760

Table 1. Simulation results of 10,000 replicates for $\lambda = 0, .5, 1, 2, 5$ and each prior.

Table 1 summarizes simulation results of 10,000 replicates for each prior. For all three priors, we observe similar tendencies. First, $E(\pi_{\hat{\theta},2})$ gets closer to $\theta = .2$ as λ increases. Second, $V(\pi_{\hat{\theta},2})$ decreases as λ decreases to zero. The average square distance between $\pi_{\hat{\theta},2}$ and $\gamma = .2$ measures a balance between $|E(\pi_{\hat{\theta},2}) - .2|$ and $V(\pi_{\hat{\theta},2})$, and the superiority depends on priors. Lastly, as $\lambda \rightarrow 0$, we have larger $P(3 \leq \sum_{i=1}^{20} Y_i \leq 5)$ and more robust $E(\sum_{i=1}^{20} Y_i)$ to prior elicitation.

In summary, when we emphasize more on population-level ethics, we have a smaller variance in the estimation for future patients (with a greater absolute bias, potentially due to Jensen’s Inequality), and the distribution of $\sum_{i=1}^n Y_i$ becomes more robust to prior elicitations. When we emphasize more on individual-level ethics, we have a larger variance in the estimation, and the distribution of $\sum_{i=1}^n Y_i$ becomes more sensitive to prior elicitations.

5. Consensus prior

In Bayesian inference, researchers are able to utilize information, which is independent of observed data. It allows researchers to incorporate any form of information, such as one’s experience and existing literature, which may be particularly useful in a small-sample study. On the

other hand, we concern subjectivity and prior sensitivity in sparse data. Furthermore, it is possible to have disagreement among multiple researchers' prior elicitations about a parameter θ .

Suppose there are K researchers with their own prior density functions, say $f(\theta|Q_k)$ for $k = 1, \dots, K$, and they have the same likelihood function $f(\vec{y}|\theta)$. Each prior elicitation leads to a unique Bayes estimator

$$\hat{\theta}_k = E(\theta|\vec{y}, Q_k) = \int \theta f(\theta|\vec{y}, Q_k) d\theta, \quad (33)$$

where $f(\theta|\vec{y}, Q_k) \propto f(\vec{y}|\theta)f(\theta|Q_k)$ is the posterior density function of θ given data \vec{y} and the k^{th} prior elicitation Q_k . For posterior estimation, one reasonable approach to compromise is a weighted average $\sum_{k=1}^K w_k \hat{\theta}_k$, where $w_k > 0$ for $k = 1, \dots, K$ and $\sum_{k=1}^K w_k = 1$. In this section, we discuss two different weighting methods. The first method is to fix w_k before observing data (referred to as prior weighting scheme). The second method is to determine $w_k(\vec{y})$ after observing data \vec{y} so that $w_k(\vec{y})$ increases when the k^{th} prior elicitation Q_k is better supported by the observed data \vec{y} (referred to as posterior weighting scheme) [5].

For a prior weighting scheme, we denote $w_k = P(Q_k)$ which quantifies the credibility of the k^{th} prior elicitation. For a posterior weighting scheme, we consider

$$w_k(\vec{y}) = P(Q_k|\vec{y}) = \frac{f(\vec{y}|Q_k)P(Q_k)}{\sum_{j=1}^K f(\vec{y}|Q_j)P(Q_j)} = \frac{w_k f(\vec{y}|Q_k)}{\sum_{j=1}^K w_j f(\vec{y}|Q_j)}, \quad (34)$$

where $f(\vec{y}|Q_k) = \int f(\vec{y}|\theta)f(\theta|Q_k) d\theta$ is the marginal likelihood from the k^{th} prior elicitation.

This formulation is similar to the BMA method discussed in Section 3. It can be shown that $\sum_{k=1}^K w_k(\vec{y}) \hat{\theta}_k$ is the Bayes estimator (the posterior mean of θ) when a consensus prior $f(\theta) = \sum_{k=1}^K w_k f(\theta|Q_k)$ is used with $w_k = P(Q_k)$ [5].

Samaniego discussed self-consistency when compromised inference is used through the prior weighting scheme $\sum_{k=1}^K w_k \hat{\theta}_k$ [4]. Let θ denote a parameter of interest and

$$E(\theta) = \int \theta f(\theta) d\theta = \theta^* \quad (35)$$

be the prior expectation, the mean of the prior density function $f(\theta)$. Let $\tilde{\theta}$ denote a sufficient statistic, which serves as an unbiased estimator for θ . When we satisfy $E(\theta|\tilde{\theta} = \theta^*) = \theta^*$, it is called self-consistency [4].

Self-consistency can be achieved under simple models. For example, let $\vec{Y} = (Y_1, \dots, Y_n)$ be a random sample, where $Y_i \sim \text{Bernoulli}(\theta)$, and assume $\theta \sim \text{Beta}(a, b)$ for prior. It can be shown

that the maximum likelihood estimator $\tilde{\theta} = \frac{1}{n} \sum_{i=1}^n Y_i$ is a sufficient statistic and an unbiased estimator for θ . The posterior mean is a weighted average between θ^* and $\tilde{\theta}$ as follows

$$E(\theta|\tilde{\theta} = \theta^*) = c\theta^* + (1 - c)\tilde{\theta}, \quad (36)$$

where $c = \frac{a+b}{a+b+n}$. If we observe $\tilde{\theta} = \theta^*$, we can achieve the self-consistency because $E(\theta|\hat{\theta} = \theta^*) = \theta^*$. In words, when prior estimate and maximum likelihood estimate are identical, the posterior estimate must be consistent with the prior estimate and the maximum likelihood estimate. The self-consistency can be also achieved in the prior weighting scheme under certain conditions as illustrated in the following example.

5.1. Binomial experiment

Let $Y_i \sim \text{Bernoulli}(\pi)$ for $i = 1, \dots, n$ and assume Y_1, \dots, Y_n are independent. Suppose the k^{th} researcher specifies the prior distribution $\pi|Q_k \sim \text{Beta}(a_k, b_k)$ for $k = 1, \dots, K$. For the prior weighting scheme, let $w_k = P(Q_k)$, the prior probability for the k^{th} prior elicitation (fixed before observing data). Since $E(\pi|Q_k) = \frac{a_k}{a_k+b_k}$ and the expectation $E(\cdot)$ is a linear operator, the average of “consensus prior” is

$$E(\pi) = \int_0^1 \pi f(\pi) d\pi = \int_0^1 \pi \left(\sum_{k=1}^K f(\pi|Q_k) P(Q_k) \right) d\pi = \sum_{k=1}^K w_k \left(\int_0^1 \pi f(\pi|Q_k) d\pi \right) = \sum_{k=1}^K w_k E(\pi|Q_k). \quad (37)$$

Let $E(\pi) = \pi^*$ and suppose the K researchers observed the consistent result $\tilde{\pi} = \frac{1}{n} \sum_{i=1}^n Y_i = \pi^*$. The individual-specific Bayes estimator is as follows

$$\hat{\pi}_k = E(\pi|\tilde{\pi} = \pi^*, Q_k) = c_k E(\pi|Q_k) + (1 - c_k) \pi^*, \quad (38)$$

for the k^{th} researcher, where $c_k = \frac{a_k+b_k}{a_k+b_k+n}$. The compromised Bayes estimator is as follows

$$E(\pi|\tilde{\pi} = \pi^*) = \sum_{k=1}^K w_k \hat{\pi}_k = \sum_{k=1}^K w_k [c_k E(\pi|Q_k) + (1 - c_k) \pi^*]. \quad (39)$$

If we allow individual-specific prior elicitation a_k and b_k with the restriction $a_k + b_k = m$ for all K researchers (i.e., the same strength of prior elicitation), the value $c_k = \frac{m}{m+n}$ is constant over all researcher. By letting the constant $c_k = c$,

$$E(\pi|\tilde{\pi} = \pi^*) = c \left(\sum_{k=1}^K w_k E(\pi|Q_k) \right) + (1 - c) \pi^* \left(\sum_{k=1}^K w_k \right) = c E(\pi) + (1 - c) \pi^* = \pi^*, \quad (40)$$

so the self-consistency is satisfied.

For the posterior weighting scheme given data $\vec{y} = (y_1, \dots, y_n)$, the marginal likelihood from the k^{th} prior elicitation is as follows

$$f(\vec{y} | Q_k) = \int_0^1 f(\vec{y} | \pi) f(\pi | Q_k) d\pi = \frac{\Gamma(a_k + b_k)}{\Gamma(a_k) \Gamma(b_k)} \frac{\Gamma(a_k + s) \Gamma(b_k + n - s)}{\Gamma(a_k + b_k + n)}, \quad (41)$$

where $s = \sum_{i=1}^n y_i$ is an observed sufficient statistic. Then, the posterior weighting scheme becomes $\sum_{k=1}^K w_k(\vec{y}) \hat{\pi}_k$ with

$$w_k(\vec{y}) = \frac{w_k f(\vec{y} | Q_k)}{\sum_{j=1}^K w_j f(\vec{y} | Q_j)}, \quad (42)$$

$$\hat{\pi}_k = \frac{a_k + s}{a_k + b_k + n}.$$

If we desire an equal strength from each researcher’s prior elicitation, we may fix $a_k + b_k = m$ and $w_k = \frac{1}{K}$. In the posterior weighting scheme, it is difficult to achieve the self-consistency.

Whether self-consistency is satisfied, the practical concern is the quality of estimation such as bias, variance and mean square error. Assuming $K = 2$ researchers have disagreeing prior knowledge and a sample of size $n = 10$, let us consider three cases. Suppose two researchers express relatively mild disagreement as $(a_1, b_1) = (1, 3)$ and $(a_2, b_2) = (3, 1)$ in Case 1, relatively strong disagreement as $(a_1, b_1) = (2, 6)$ and $(a_2, b_2) = (6, 2)$ in Case 2, and even stronger disagreement as $(a_1, b_1) = (3, 9)$ and $(a_2, b_2) = (9, 3)$ in Case 3. For each case, **Figure 7** provides the relative bias, variance and mean square error (MSE) for comparing the posterior weighting scheme $\sum_{k=1}^3 w_k(\vec{y}) \hat{\pi}_k$ to the prior weighting scheme $\sum_{k=1}^3 w_k \hat{\pi}_k$. When a relative MSE is smaller than one, it implies a smaller MSE for the posterior weighting scheme. As the true value of π is well between the two prior guesses $E(\pi | Q_1) = .25$ and $E(\pi | Q_2) = .75$, the posterior weighting scheme shows a greater MSE due to greater variance. When the true value of π deviates away from either prior guess, the posterior weighting schemes show a smaller MSE due to smaller bias. The tendency is stronger when the two disagreeing prior elicitations are stronger (i.e., stronger prior disagreement). The bottom line is a clear bias-variance tradeoff

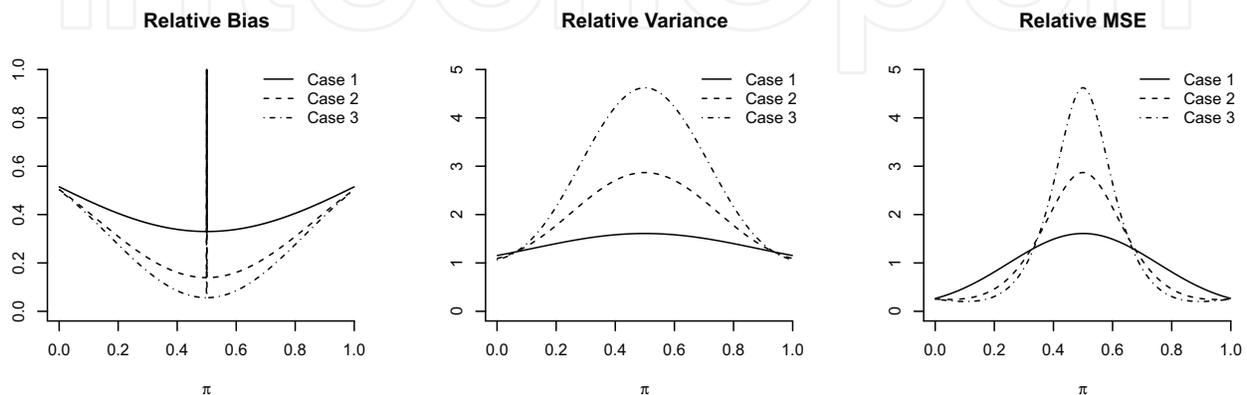


Figure 7. Comparing prior and posterior weighting schemes for different degrees of disagreements.

when we compare the two weighting schemes. $\sum_{k=1}^3 w_k(\vec{y}) \hat{\pi}_k$ is able to reduce bias when there is strong discrepancy between “consensus prior” and data, but it has larger variance than $\sum_{k=1}^3 w_k \hat{\pi}_k$ because $w_k(\vec{y})$ depends on random data.

5.2. Applications to Phase I trials under logistic regression model

In this section, we apply the prior weighting scheme and the posterior weighting scheme to Phase I clinical trials under the logistic regression model. We consider the three priors considered in Section 4.6. We denote Prior 1, 2 and 3 by Q_1 , Q_2 and Q_3 , respectively. The three priors had the same hyper-parameters $a_{-1,k} = 1, b_{-1,k} = 3, a_{0,k} = 3, b_{0,k} = 1$, but they were different by $x_{-1,k} = -4, 0, 4$ and $x_{0,k} = 4, 8, 12$ for $k = 1, 2, 3$, respectively. By the use of the conditional mean prior in Eq. (22), the prior density function of $\vec{\beta}$ for prior Q_k is given by

$$f(\vec{\beta} | Q_k) \propto (x_{0,k} - x_{-1,k}) \prod_{i=-1}^0 \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{a_{i,k}} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{b_{i,k}}. \quad (43)$$

The prior means were $E(\theta_2 | Q_1) = -1.70, E(\theta_2 | Q_2) = 1.40$ and $E(\theta_2 | Q_3) = 5.38$ for Priors 1, 2 and 3, respectively.

For simulation study, we consider three simulation scenarios with sample size $N = 20$. In Scenario 1, we assume $\beta_0 = -5$ and $\beta_1 = .6$, so the true MTD is $\theta_2 = 6.02$, which deviates significantly from all of the three prior means. In Scenario 2, we assume $\beta_0 = -3$ and $\beta_1 = .8$ as in Section 4.6, so $\theta_2 = 2.02$ is well surrounded by the three prior means. In Scenario 3, we assume $\beta_0 = -1$ and $\beta_1 = 1.2$, so $\theta_2 = -.32$ is close to the most conservative prior mean $E(\theta_2 | Q_1) = -1.70$. We consider the loss function $L_I(\vec{\beta}, x_{n+1}) = (x_{n+1} - \theta_2)^2$ discussed in Section 4.4, which focuses on individual-level ethics. We use the uniform prior probabilities $w_k = P(Q_k) = 1/3$ for $k = 1, 2, 3$ for implementing both prior and posterior weighting scheme.

Table 2 provides the simulation results of 10,000 replicates for each scenario under the prior weighting scheme and under the posterior weighting scheme. Since the posterior weighting scheme adaptively updates $w_k(\vec{y})$ based on empirical evidence, it can reduce bias, but it has greater variance in the estimation of θ_2 . As a consequence, when the true MTD was close to one extreme prior estimate (Scenarios 1 and 3), the use of the posterior weighting scheme yields a smaller $E[(\pi_{\hat{\theta}_2} - .2)^2]$, $E(\sum_{i=1}^{20} Y_i)$ closer to $N\gamma = 4$, and $P(3 \leq \sum_{i=1}^{20} Y_i \leq 5)$ closer to one when compared to the use of the prior weighting scheme. In Scenario 3, the average number of adverse events was 4.6 for the posterior weighting scheme, but it was as high as 7.1 in the prior weighting scheme. On the other hand, when the true MTD was well surrounded by the three prior estimates (Scenario 2), the use of the prior weighting scheme yielded more plausible results.

The simulation results are analogous to the simpler model in Section 5.1. When the true parameter is not well surrounded by prior guesses, the posterior weighting scheme is preferable with respect to mean square error due to smaller bias. When the true parameter is well surrounded by prior guesses, the prior weighting scheme is beneficial with respect to mean square error due to smaller variance.

Scenario	Method	$E(\pi_{\hat{\theta}_2})$	$V(\pi_{\hat{\theta}_2})$	$E[(\pi_{\hat{\theta}_2} - .2)^2]$	$E(\sum_{i=1}^{20} Y_i)$	$P(3 \leq \sum_{i=1}^{20} Y_i \leq 5)$
1	Prior weighting	0.0967	0.0014	0.0121	1.1090	0.0398
	Posterior weighting	0.1853	0.0073	0.0075	2.7304	0.5900
2	Prior weighting	0.2018	0.0059	0.0059	3.8432	0.9042
	Posterior weighting	0.2048	0.0110	0.0110	4.2848	0.8920
3	Prior weighting	0.2929	0.0071	0.0157	7.1090	0.0568
	Posterior weighting	0.1951	0.0133	0.0133	4.6036	0.8646

Table 2. Simulation results of 10,000 replicates for the prior and posterior weighting schemes.

As a final comment, we shall be careful about the strength of individual prior elicitations when we implement the posterior weighting scheme in Phase I clinical trials. The strength of individual prior elicitation depends on (i) the hyper-parameters $a_{i,k}$ and $b_{i,k}$, (ii) the prior weight $w_k = P(Q_k)$ as well as (iii) the distance between the two arbitrarily chosen doses $x_{0,k} - x_{1,k}$. It can be seen through the expression

$$f(\vec{\beta}) = \sum_{k=1}^K f(\vec{\beta}|Q_k) P(Q_k) \propto \sum_{k=1}^K w_k (x_{0,k} - x_{1,k}) \prod_{i=-1}^0 \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{a_{i,k}} \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right)^{b_{i,k}}. \quad (44)$$

When researchers determine consensus prior elicitation before initiating a trial, the multiplicative term $w_k (x_{0,k} - x_{1,k})$ shall be carefully considered together with the hyper-parameters $a_{i,k}$ and $b_{i,k}$ [5].

6. Concluding remarks

In this chapter, we have discussed Bayesian inference with averaging, balancing, and compromising in sparse data. In the cancer risk assessment, we have observed that low-dose inference can be very sensitive to an assumed parametric model (Section 3.1). In this case, the Bayesian model averaging can be a useful method. It provides robustness by using multiple models and posterior model probabilities to account for model uncertainty. In the application of Bayesian decision theory to Phase I clinical trials, we have observed that the sequential sampling scheme heavily depends on a loss function. A loss function, which is devised from individual-level ethics, focuses on the benefit of trial participants, and a loss function, which is devised from population-level ethics, focuses on the benefit of future patients. It is possible to balance between the two conflicting perspectives, and we can adjust a focusing point by the tuning parameter (Sections 4.5 and 4.6). Finally, the use of a weighted posterior estimate can be a compromising method when two or more researchers have prior disagreement. We have compared the prior and posterior weighting schemes in a small-sample binomial problem (Section 5.1) and in a small-sample Phase I clinical trial (Section 5.2). The prior weighting scheme (data-independent weights) outperforms when prior estimates surround the truth, and the posterior weighting scheme (data-dependent weights) outperforms when the truth is not well surrounded by prior estimates. One method does not outperform the other method for all parameter values, so it is important to be aware of their bias-variance tradeoff.

Author details

Steven B. Kim

Address all correspondence to: stkim@csumb.edu

Department of Mathematics and Statistics, California State University, CA, United States

References

- [1] Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*. 1997;**92**:171-191
- [2] Whitehead J, Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*. 1998;**8**:445-467
- [3] Kim SB, Gillen DL. A Bayesian adaptive dose-finding algorithm for balancing individual- and population-level ethics in Phase I clinical trials, *Sequential Analysis*. 2016;**35**(4):423-439
- [4] Samaniego FJ. *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. New York: Springer; 2010
- [5] Kim SB, Gillen DL. An alternative perspective on consensus priors with applications to Phase I clinical trials. *Jacobs Journal of Biostatistics*. 2016;**1**(1):006
- [6] Shao K, Small MJ. Potential uncertainty reduction in model-averaged benchmark dose estimates informed by an additional dose study. *Risk Analysis*. 2011;**31**:1156-1175
- [7] Bedrick EJ, Christensen R, Johnson W. A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*. 1996;**91**(436):1450-1460
- [8] International Agency for Research on Cancer. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Vol. 69. Lyon: IARC; 1997. ISBN 92-832-1269-X
- [9] Kociba RJ, Keyes DG, Beyer JE, Carreon RM, Wade CE, Dittenber DA. et al. Results of a two-year chronic toxicity and oncogenicity study of 2,3,7,8-tetrachlorodibenzo-p-dioxin in rats. *Toxicology and Applied Pharmacology*. 1978;**46**(2):279-303
- [10] Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Statistical Science*. 1999;**14**(4):382-417
- [11] Simmons SJ, Chen C, Li X, Wang Y, Piegorsch WW, Fang Q, Hu B, Dunn GE. Bayesian model averaging for benchmark dose estimation. *Environmental and Ecological Statistics*. 2015;**22**(1):5-16
- [12] Crump KS. A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology*. 1984;**4**:854-871
- [13] EPA (US Environmental Protection Agency). *Benchmark dose technical guidance, EPA/100/R-12/001, Risk Assessment Forum*. Washington, DC: U.S. Environmental Protection Agency; 2012

- [14] Whitehead J, Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*. 1998;**8**:445-467
- [15] O'Quigley J, Conaway M. Continual reassessment and related dose-finding designs. *Statistical Science*. 2010;**25**:202-216
- [16] Bartroff J, Lai TL. Incorporating individual and collective ethics into Phase I cancer trial designs. *Biometrics*. 2011;**67**:596-603
- [17] O'Quigley J, Pepe M, Fisher L. Continual reassessment method: A practical design for Phase 1 clinical trials in cancer. *Biometrics*. 1990;**46**:33-48
- [18] Whitehead J, Brunier H. Continual reassessment method: Bayesian decision procedures for dose determining experiments. *Statistics in Medicine*. 1995;**14**:885-893