Hierarchical Bayesian Benchmark Dose Analysis

by
Qijun Fang

A Dissertation Submitted to the Faculty of the

Graduate Interdisciplinary Program in Statistics

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 1 4

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Qijun Fang entitled
Hierarchical Bayesian Benchmark Dose Analysis
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: Feb 4, 2014
Walter W. Piegorsch

_____ Date: Feb 4, 2014
Rabindra N. Bhattacharya

_____ Date: Feb 4, 2014
D. Dean Billheimer

_____ Date: Feb 4, 2014
Chengcheng Hu

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: Feb 4, 2014
Dissertation Director: Walter W. Piegorsch

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: Qijun Fang

Acknowledgments

First and foremost, thanks go to my advisor Dr. Walter Piegorsch. For what I have learned in wide areas of Statistics through him, his excellent guidance in this project as well as his constant encouragement and support throughout the years. His creative ideas and the time we spent on discussion were absolutely essential to the completion of this project.

Thanks are due my dissertation committee members, Drs. Katherine Y. Barnes, Rabi Bhattacharya, Anton Westveld, D. Dean Billheimer and Chengcheng Hu for their precious suggestions during my research on this project. Special thanks are due Dr. Rabi Bhattacharya for giving me wonderful lectures in advanced theoretic statistics and offering me many other great helps during my doctoral study. Also, thanks are due my colleagues in the University of North Carolina at Wilmington, Drs. Susan J. Simmons, Cuixian (Tracy) Chen and Yishi Wang for their helpful comments and suggestions during the preparation of the publications regarding to this research. I also want to thank all who have taught me mathematics and statistics through those interesting courses.

Further thanks are due my supervisor for my internship at the Biostatistis and Data Management department (BADM) of the Ventana Medical System, Inc, Dr. James Ranger-Moore for providing me funding for my doctoral study, broadening my view of statistical applications in medical industry and allowing me to use the BADM workstation for computational intensive tasks in this project. Special thanks are due my manager Dr. Isaac Bai for teaching me to use SAS software and sharing with me his experience working in industry. I would also like to thank all my other colleagues in Ventana for it is so pleasant to work with them during the past five years.

Finally I would like to thank my wife Fang for her love and unconditional support.

Dedication

To my wife, my parents, my grandparents and my arriving baby.

Table of Contents

TABLE OF CONTENTS—*Continued*

List of Figures

List of Figures—*Continued*

LIST OF FIGURES—*Continued*

List of Tables

List of Tables—*Continued*

Abstract

An important objective in statistical risk assessment is estimation of minimum exposure levels, called Benchmark Doses (BMDs) that induce a pre-specified Benchmark Response (BMR) in a target population. Established inferential approaches for BMD analysis typically involve one-sided, frequentist confidence limits, leading in practice to what are called Benchmark Dose Lower Limits (BMDLs). Appeal to hierarchical Bayesian modeling and credible limits for building BMDLs is far less developed, however. Indeed, for the few existing forms of Bayesian BMDs, informative prior information is seldom incorporated. Here, a new method is developed by using reparameterized quantal-response models that explicitly describe the BMD as a target parameter. This potentially improves the BMD/BMDL estimation by combining elicited prior belief with the observed data in the Bayesian hierarchy. Besides this, the large variety of candidate quantal-response models available for applying these methods, however, lead to questions of model adequacy and uncertainty. Facing this issue, the Bayesian estimation technique here is further enhanced by applying Bayesian model averaging to produce point estimates and (lower) credible bounds. Implementation is facilitated via a Monte Carlo-based adaptive Metropolis (AM) algorithm to approximate the posterior distribution. Performance of the method is evaluated via a simulation study. An example from carcinogenicity testing illustrates the calculations.

Major Professor: Walter W. Piegorsch

Chapter 1
Introduction

1.1 Quantitative Risk Assessment

Quantitative risk assessment involves estimation of the severity and likelihood of adverse responses associated with exposure to hazardous stimuli. With this context, one can, e.g., assess many forms of risk such as biological disease risks (Wakefield, 2008), environmental health risks (Stern, 2008), ecological risks (Fox, 2006), economic loss due to natural disasters (Boruff *et al.*, 2006), industrial or engineering risks such as nuclear power plant failures (Grimston, 2002), etc. This dissertation will exclusively focus on risks to biological (including human) or ecological systems. In these situations, 'risk' can be defined as follows.

Definition 1.1.1. *Risk, R(d), is defined as the probability of some pre-defined adverse effect, such as death, weight loss, birth defect, cancer or mutation exhibited in a subject exposed to a particular dose level, d, of a hazardous agent.*

Definition 1.1.1 defines risk in terms of an adverse effect. Thus, it is reasonable to assume that $R(d)$ is a monotone increasing function of dose. Although seemingly straightforward, this definition contains an important, implicit feature: non-zero risk may exist, even change, for very small levels of d. This extends earlier concepts of risk where, at least for many non-cancer endpoints, one assumed that some dose threshold exists below which $R(d) = 0$. By modeling the risk more formally, however, a richer variety of possible dose-response functions and consequent statistical machinery becomes available (Krewski and van Ryzin, 1981). Notice that the unknown quantity $R(0)$ represents the risk to which all subjects in a population are exposed. To correct for this spontaneous risk of response, additional risk and extra risk functions

frequently are used in risk estimation (Piegorsch and Bailer, 2005, §4.2.1). Defined formally, these are as follows:

Definition 1.1.2. *Additional risk, $R_A(d)$, is the risk beyond that of the control (background, spontaneous) level; that is:*

$$R_A(d) = R(d) - R(0). \tag{1.1.1}$$

Definition 1.1.3. *Extra risk, $R_E(d)$, is the additional risk among subjects who on average would not have responded under control conditions:*

$$R_E(d) = \frac{R_A(d)}{1 - R(0)} = \frac{R(d) - R(0)}{1 - R(0)}, \text{ where } R(0) < 1. \tag{1.1.2}$$

Notice that when the background risk is 0, i.e., $R(0) = 0$, the additional risk and extra risk are both equal to $R(d)$. Moreover, if $0 \leq R(0) < 1$ ($R(0)$ seldom equals 1 but it may equal 0), then $R_E(d) \geq R_A(d)$; in other words, extra risk can be thought of as the additional risk after inflation to account for control-condition non-response. Together, $R_E(d)$ and $R_A(d)$ are known as forms of *excess risk*.

Quantitative risk assessment in public health or environmental applications is usually broken down into four fundamental steps (Stern, 2002). First is *hazard identification*, where an agent, substance, or other environmental stimulus induces detrimental outcomes in some industrial, occupational, public health, or ecological setting. The detrimental outcomes envisioned in this dissertation may be cancer, mutation, birth defect, ecosystem damage, etc. Second is *stimulus/dose-response assessment*, where any quantifiable relationship between the agent and the detrimental outcome is modeled and estimated. Here the term *dose* is used as a generic label for any quantification of a hazardous exposure. Third is *exposure assessment*, where the extent, frequency, and duration of the exposure before and after application of regulatory measures are determined. Last is *risk characterization* in which the risk analyst incorporates information from the previous three stages into a single assessment of the

overall risk due to exposure to the hazardous agent. Using these four steps as a basic paradigm for the risk assessment process, it is assumed herein that identification of the hazard has been addressed and that the focus will be on dose-response assessment and quantitative risk estimation.

1.2 Low-Dose Risk Estimation and Benchmark Risk Analysis

In many cases, epidemiological or human data on hazardous substances are not available or are inadequate for quantitative risk assessment. This may be due to the lack of accurate information on exposure levels or the confounding of risk factors where it is impossible to separate the effects of one hazard from another (Krewski *et al.*, 1991; Bailer and Portier, 1994). Thus, to assess the adverse effects of a hazardous agent, bioassays are often conducted on laboratory rodents (mice, rats, etc.) or other biological systems such as aquatic animals or cells in a laboratory culture. Due to the short life span of laboratory animals and to guarantee that a toxic effect will be observable within a reasonably short period of time (say, several weeks or months), the dose levels of the agent(s) are administered at relatively high values. This is true primarily for laboratory animal experiments conducted as screens for certain toxic effects (Haseman, 1984, 1985; Piegorsch, 1994). Unfortunately, the dose-response pattern exhibited at high doses in such a study may not apply in the low-dose region. Many candidate dose-response models may fit data equally well at high doses, but may also yield dramatically different estimates at lower does levels (Piegorsch and Bailer, 2005, §4.2.2). The so-called *low-dose extrapolation* problem for estimating risk from high dose levels to lower doses of regulatory interest (Brown and Koziol, 1983) is one of the greatest challenges in quantitative assessment of possible human or ecological risks.

A contemporary approach to low-dose estimation, known as *benchmark risk analysis*, uses the functional specification for $R(d)$ to provide low-dose estimates for risk

and/or excess risk. In this approach, the *benchmark dose* at which a predetermined level of risk is attained is estimated by inverting the dose-response relationship. Formal definitions used in benchmark analysis are as follows:

Definition 1.2.1. *The Benchmark Dose (BMD) is an exposure due to a dose of a substance associated with a specified low incidence of risk, generally in the range of* 1% *to* 10%*, of a health effect; or the dose associated with a specified measure or change of a biological effect.*

Definition 1.2.2. *The Benchmark Response (BMR) is a response, generally expressed as excess of background (e.g., extra risk), at which a benchmark dose is desired.*

For a typical data setting (for example, the quantal data which will be introduced in §1.3), the BMD is usually found by solving for the smallest $d \geq 0$ that satisfies $R_E(d)$=BMR.

Definition 1.2.3. *The Benchmark Dose Lower Limit (BMDL) is a lower one-sided 1-α confidence (or credible) limit on the BMD.*

It is common to use the BMDL as a point of departure which is then reduced by a set of uncertainty/safety factors to arrive at an acceptable level of human exposure or to otherwise establish human low-exposure guidelines (Gaylor, 1998). BMDLs corresponding to BMR = 0.01, 0.05 or 0.10 are most often seen in practice for a given adverse effect (U.S. EPA, 2012). Where needed for clarity, a subscript is added for the BMR level at which each quantity is calculated: $\text{BMD}_{100\text{BMR}}$ and $\text{BMDL}_{100\text{BMR}}$. In this fashion, use of BMDs and BMDLs for quantifying and managing risk is growing in both the United States and the European Union (U.S. General Accounting Office, 2001; European Union, 2003; OECD, 2006, 2008).

1.3 Quantal Response Data and Binomial Framework

In risk-analytic dose-response studies, 'quantal' data are often observed. With this type of data, the observations are in the form of proportions and the experimental subjects are classified in a binary fashion as either exhibiting or not exhibiting the adverse effect. Such data are very common in carcinogenesis, teratogenesis and mutagenesis studies (Crump et al., 1976; Chen and Kodell, 1989; Gaylor, 1989, 1998).

With quantal data, benchmark analysis is usually performed under a binomial framework. Denote Y_i as the number of responses at the i^{th} dose level, out of N_i subjects tested at that dose ($i = 1, ..., m$). The standard statistical model assumes $Y_i \sim$ indep. $Binom\big(N_i, R(d_i)\big)$, where $R(d_i)$ is the risk at dose d_i. For generic purposes, denote $\boldsymbol{\theta}$ as an unknown parameter vector that describes $R(d)$. Under the independence assumption, the joint probability mass function (p.m.f.) for $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ becomes

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{m} f(Y_i|\boldsymbol{\theta}) = \prod_{i=1}^{m} \binom{N_i}{Y_i} R(d_i)^{Y_i}\{1 - R(d_i)\}^{N_i - Y_i}, \tag{1.3.1}$$

where $f(Y_i|\boldsymbol{\theta})$ is the individual binomial p.m.f. for each Y_i.

Traditionally, estimation of the BMD has been performed via maximum likelihood: viewing the joint p.m.f. as a likelihood function, $L(\boldsymbol{\theta}) = f(\boldsymbol{Y}|\boldsymbol{\theta})$, maximizing $L(\boldsymbol{\theta})$ or $\log\{L(\boldsymbol{\theta})\}$ produces maximum likelihood estimators (MLEs) $\widehat{\boldsymbol{\theta}}$ for the parameter vector $\boldsymbol{\theta}$. The corresponding MLEs $\widehat{R}(d)$ for $R(d)$ and $\widehat{R}_E(d)$ for $R_E(d)$ are then obtained using the invariance property of MLE. Setting $\widehat{R}_E(d) =$ BMR and solving for d yields the MLE, $\widehat{\text{BMD}}_{100\text{BMR}}$, for the target quantity, BMD. This method is a form of inverse non-linear regression and, except for the use of an excess risk function upon which to base the inversion, is essentially equivalent to estimation of an 'effective dose' such as the well-known median effective dose, ED_{50} (Piegorsch and Bailer, 2005, §4.1.1). The corresponding BMDL is then built from the statistical properties of $\widehat{\text{BMD}}$; parametric possibilities include Wald-type lower confidence limits (Moerbeek

et al., 2004), profile likelihood limits (Crump and Howe, 1985), or appeal to the bootstrap (West *et al.*, 2009). The latter has also been employed for constructing non-parametric BMDLs (Piegorsch *et al.*, 2012).

1.4 Hierarchical Bayesian Modeling

The maximum likelihood approach mentioned above for calculating BMDs typically involves non-hierarchical models and in particular, frequentist confidence limits for the BMDL. Different from the frequentist framework, Bayesian methods view this problem in a hierarchical perspective. Here, $\boldsymbol{Y}$ is modeled conditionally on $\boldsymbol{\theta}$. If $\boldsymbol{\theta}$ is itself thought to be random, say, $\pi(\boldsymbol{\theta})$, application of Bayes rule (Casella and Berger, 2002, §1.3) results in a form of *updated* information on $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{\pi(\boldsymbol{\theta})f(\boldsymbol{Y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta})f(\boldsymbol{Y}|\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{1.4.1}$$

where $f(\boldsymbol{Y}|\boldsymbol{\theta})$ (known as the *likelihood function*) represents the parent distribution indexed by $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$ is the *prior density function* for $\boldsymbol{\theta}$ (the distribution corresponding to it is known as the *prior distribution*), and $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ is called the *posterior density function* for $\boldsymbol{\theta}$ after updating with the data $\boldsymbol{Y}$ (the distribution corresponding to this density is known as the *posterior distribution*).

This hierarchical structure for the data and $\boldsymbol{\theta}$ is often viewed from what is known as the *Bayesian formulation* for statistical inference (Casella and Berger, 2002, §§7.2.3, 9.2.4). Simply put, a fully specified posterior distribution for $\boldsymbol{\theta}$ tells a Bayesian all there is to know about $\boldsymbol{\theta}$; estimation and inferences follow directly from this.

1.5 Established Bayesian Benchmark Dose Analysis

For estimating a benchmark dose, the BMD is viewed as one of the unknown parameters in $\boldsymbol{\theta}$, along with any other nuisance parameters required by the posited

dose-response model. For simplicity, denote ξ as BMD. Hereupon, emphasis will be restricted to fully parametric Bayesian estimation for and inference on ξ.

Established Bayesian benchmark dose analysis methods typically use generalized linear models (GLiM) to model the dose-response relationship under a binomial likelihood. In most models, ξ is not included in the default parameter set; instead, $R(d)$ is usually modeled via a linear predictor:

$$g(R(d)) = \beta_0 + \beta_1 x_1(d) + \beta_2 x_2(d) + \ldots, \tag{1.5.1}$$

where $g(\cdot)$ is called the *link function* which maps the probability value to the whole real line and $x_i(d)$ are functions of d which are monotone on $[0, \infty)$ (Brown and Prescott, 2006, §3.1.1). Fully parametric Bayesian benchmark risk analyses with quantal data which have appeared in the literature have employed this form (or variants thereof). The parameters in these models usually don't have pertinent risk analytic interpretations, hence it is difficult to construct pertinent prior distributions for them. As a consequence, the prior distributions for these parameters are often chosen to be 'non-informative' or 'improper'. (An improper prior has an infinite integral over the range $(-\infty, \infty)$; a non-informative prior is any prior with a very flat density curve, loosely speaking, but the integral over the range $(-\infty, \infty)$ is finite; see Gelman *et al.* (2004, §2.9) and Geweke (2005, §3.2).) The corresponding hierarchical models lose the ability to intelligently incorporate the analyst's prior knowledge and the inferences based on these priors rely heavily on the input data (Gelman *et al.*, 2004, §2.9). For example, Shao and Small (2011) use the well-known logistic model $R(d) = \{1 + \exp(-\beta_0 - \beta_1 d)\}^{-1}$ and the so-called quantal-linear model $R(d) = 1 - \exp(-\beta_0 - \beta_1 d)$. The latter is a special case of the more general multi-stage (k-stage) model $R(d) = 1 - \exp(-\beta_0 - \beta_1 d - \ldots - \beta_k d^k)$. They apply essentially non-informative normal priors for β_0 and β_1 for modeling quantal data in a tumorigenicity experiment; further developments appeared in Shao and Small (2012). Shao (2012) expanded these considerations to the probit model $R(d) = \Phi(\beta_0 + \beta_1 d)$, where $\Phi(\cdot)$ is

the CDF for standard normal distribution, and also introduced a power prior to build historical control information into the hierarchy. (A variety of quantal response models, including the logistic, probit and quantal-linear will be introduced in §2.1, below.) Wang *et al.* (2011) use the logistic model and the probit model with non-informative hierarchical normal, gamma and uniform priors for the β-parameters to study BMDs for nephrolithiasis in children. Naufal *et al.* (2009) use the entire suite of dose-response models in the EPA's Benchmark Dose software, BMDS (Davis *et al.*, 2012), except for its quantal-quadratic model (see Table 2.1, below), with non-informative uniform priors for the β-parameters to study the effect of a tobacco specific nitrosamine (NNK) on oral cancer. Other fully parametric Bayesian benchmark risk analyses with continuous or count data are also based on forms of generalized linear models. For example, Shao and Gift (2014) apply non-informative priors for the β-parameters in four dose-response models for continuous data listed in BMDS to study model uncertainty issue in continuous data benchmark dose analysis. Wheeler and Bailer (2009a) use Poisson regression models with non-informative normal priors for the mean parameters and non-informative inverse gamma priors for the variance parameters to study the effect of NaCl on the reproducibility of *Ceriodaphnia dubia* (a water flea which is used in toxicity testing of waste-water treatment plant effluent water in the United States). Morales *et al.* (2006) use similar models with non-informative normal priors for the parameters to study the effect of arsenic in drinking water on lung cancer and bladder cancer. Held (2004) uses a quadratic regression model with improper priors for the β-parameters to model the effects of copper toxicity in sea kelp.

Similar to the BMD, an approach to inverse dose estimation is known as the *effective dose* ρ. This is defined as the dose that yields a $100\rho\%$ effect over the quantal dose-response curve, for $0 < \rho < 1$ (Piegorsch and Bailer, 2005, §4.1.2). Common notation for this is $\mathrm{ED}_{100\rho}$. Effective dose analysis is similar to benchmark risk analysis in that both methods invert a function to estimate a dose that quantifies a critical effect. The effective dose is obtained by inverting the risk function, i.e., solving $R(d) =$

ρ. By contrast, the BMD as employed here with quantal data solves $R_E(d) = \text{BMR}$. For finding a Bayesian $\text{ED}_{100\rho}$, Chen (2010) uses the logistic model with improper priors for the unknown β-parameters. Li *et al.* (2008) use the logistic model with three types of priors (improper, exponential and conjugate) for the unknown β-parameters. Hu *et al.* (2008) use the logistic model with both improper priors and uniform priors for the β-parameters. Sun and Tsutakawa (1997) consider a one-parameter and a two-parameter logistic model with non-informative normal priors and gamma priors for the parameters to derive a Bayesian design in order to minimize the expected posterior variances for the $\text{ED}_{100\rho}$. Kuo and Cohen (1999) use multi-stage models (see above, with number of stages equaling 1, 2 and 3) with improper priors for the parameters to estimate a Bayesian effective dose. Notably, Hu *et al.* (2008) and Sun and Tsutakawa (1997) reparameterize their models to include the $\text{ED}_{100\rho}$ as a parameter in the model. This provides an opportunity to perform direct inferences through the posterior distribution on the effective dose.

In the above examples, Bayesian estimates of the BMD or ED are usually obtained indirectly. A necessary first step is usually to estimate the traditional model parameters such as β_0 and β_1 in (1.5.1). Posterior density approximation such as the Laplace's method (Zellner and Rossi, 1984) or posterior simulation such as the Markov chain Monte Carlo (McMC) methods (Barnes, 2012) are most popular tools to obtain these Bayesian estimates. For instance, Shao and Small (2011, 2012); Shao (2012); Shao and Gift (2014) first obtain McMC samples of the β-parameters and then substitute these samples of parameters into the expression for BMD to obtain a sample of estimated BMDs. Then, they use the median or the mean of the McMC sample of estimated BMDs as potential alternatives for their Bayesian estimate. Wang *et al.* (2011) first use McMC to obtain the Bayesian estimate of the β-parameters in the model, and then substitute these parameter estimates into their expression for BMD to obtain the estimate of BMD. Wheeler and Bailer (2009a) first find the posterior mean of the β-parameters through McMC, and then substitute these Bayesian esti-

mates of the parameters into their expression for BMD to find the Bayes estimate of BMD. Faes et al. (2006) find a Bayesian estimate of BMD by inverting the posterior mean of the estimated extra risk function, whose parameters are estimated via the Gibbs sampler (a special case of the McMC method). Morales *et al.* (2006) first estimate the relative risk associated with the exposure concentration, and then substitute to find the estimated BMD. Held (2004) first obtains Monte Carlo samples of the additional risk, and then obtains the posterior median of the additional risk, after which he finds the estimated BMD by inverting the posterior median of the additional risk. Chen (2010) first finds estimates of the parameters via Laplace's method and iterative reweighted least square (IRLS) and then substitutes into his expression for $\mathrm{ED}_{100\rho}$ to find the estimate of ED. Kuo and Cohen (1999) first obtain the posterior mean of the parameters and then substitute to find the estimated ED.

In all these approaches, a BMDL or a credible interval for the ED is often obtained by first constructing a Monte Carlo sample of the estimated BMD or ED and then taking, for example, the lower 5^{th} percentile of the sample as the 95% lower limit; see, e.g., Shao and Small (2011, 2012), Shao (2012), Naufal *et al.* (2009), Wheeler and Bailer (2009a), Morales *et al.* (2006) and Kuo and Cohen (1999). Some authors also obtain the BMDL by first obtaining an upper credible limit of the excess risk function and then inverting this to find the BMDL; see, e.g., Faes et al. (2006) and Held (2004). Other methods to find a BMDL include use of Fieller's theorem (Chen, 2010), obtaining an approximate bivariate distribution of the BMD via a two-dimensional smoothing kernel and then obtaining the contour (Li *et al.*, 2008) or using adaptive direction sampling and Hyndman's method (Hu *et al.*, 2008).

Hierarchical Bayesian modeling in quantitative risk assessment may also consider multiple responses or multiple agents. Faes et al. (2006) jointly consider the probability of fetal death, fetal malformation and fetal weight loss in studying the impact of Ethylene Glycol (EG) on the fetus. Li *et al.* (2008) and Hu *et al.* (2008) consider more than one agent in their respective studies. Besides fully parametric Bayesian

methods, semi-parametric or non-parametric methods are also seen. For example, Wheeler and Bailer (2012) considered semi-parametric models for the dose response, incorporating a probit kernel and cubic B-splines. They placed normal priors on the basis-function coefficients and built dose-response monotonicity into their prior hierarchical constraints. Noteworthily, Wheeler and Bailer built informative priors at $d = 0$ to incorporate potential historical control information. Guha *et al.* (2013) described a nonparametric Bayesian model for the quantal setting, with beta/Dirichlet priors on pertinent probabilities related to their nonparametric construction. Hierarchical Bayesian modeling is also used to perform model averaging (a method to obtain a weighted-averaged estimator across a variety of models). For example, Shao and Small (2011, 2012); Shao (2012); Shao and Gift (2014) and Morales *et al.* (2006) perform fully Bayesian model averaging (BMA) to obtain an averaged estimate of the BMD. Details for BMA will be discussed in Chapter 4.

The goal of the dissertation is to obtain an improved estimation and calculation archetype for the BMD and for the BMDL, by applying various Bayesian strategies for incorporating direct prior information about the BMD via a reparameterized binomial likelihood. Chapter 2 describes a suite of models and the reparameterizations for them. Chapter 3 develops a Bayesian framework using the reparameterized $R(d)$'s. Chapter 4 discusses the issues of model adequacy/uncertainty and describes Bayesian model averaging that attempts to provide a more model-robust option for BMD/BMDL estimation. Chapter 5 follows with a real carcinogenicity data example. Chapter 6 examines the characteristics of the Bayesian methodology and the performance of Bayesian model averaged BMDL estimates via a series of simulation studies. Chapter 7 concludes the dissertation.

Chapter 2

Reparameterization of Quantal-Response Models

2.1 Suite of Quantal-Response Models

Focus here is on the quantal data setting, with a binomial likelihood. Previous parametric presentations for modeling $R(d)$ have focused on a suite of eight different functions (Naufal *et al.*, 2009; Wheeler and Bailer, 2009b; Piegorsch *et al.*, 2013), corresponding to popular choices in the U.S. EPA's BMDS software (Davis *et al.*, 2012). These eight models, as well as any constraints or bounds on the parameters, are reproduced in Table 2.1.

Table 2.1. Risk functions for common quantal-response models

Code	**Name**	**Risk Function $R(d)$**	**Constraints**
M1	Logistic	$\frac{1}{1+\exp(-\beta_0-\beta_1 d)}$	
M2	Probit	$\Phi(\beta_0+\beta_1 d)$	
M3	Quantal-Linear	$1-\exp(-\beta_0-\beta_1 d)$	$\beta_0 \geq 0,\ \beta_1 \geq 0$
M4	Quantal-Quadratic	$\gamma_0+(1-\gamma_0)(1-\exp(-\beta_1 d^2))$	$0 \leq \gamma_0 < 1,\ \beta_1 \geq 0$
M5	Two-Stage	$1-\exp(-\beta_0-\beta_1 d-\beta_2 d^2)$	$\beta_0, \beta_1, \beta_2 \geq 0$
M6	Log-Logistic	$\gamma_0+\frac{1-\gamma_0}{1+\exp(-\beta_0-\beta_1 \ln(d))}$	$0 \leq \gamma_0 < 1,\ \beta_1 \geq 0$
M7	Log-Probit	$\gamma_0+(1-\gamma_0)\Phi(\beta_0+\beta_1 \ln(d))$	$0 \leq \gamma_0 < 1,\ \beta_1 \geq 0$
M8	Weibull	$\gamma_0+(1-\gamma_0)(1-\exp(-e^{\beta_0} d^{\beta_1}))$	$0 \leq \gamma_0 < 1,\ \beta_1 \geq 1$

Notice in the table that models M_1–M_4 employ only two unknown parameters, while models M_5–M_8 employ three.

From the risk functions in Table 2.1, the extra risk functions (for definition, see Chapter 1) for each model are derived. As mentioned earlier, the BMD is the smallest positive solution when solving for d in the equation constructed by setting the extra risk function to equal the BMR. Recall that BMD is denoted as ξ. The extra risk functions and the corresponding ξ's for all eight models in Table 2.1 are shown in Table 2.2:

TABLE 2.2. Extra Risk functions and BMDs for quantal-response models. Note: BMR$\in (0,1)$ is the benchmark response and BMD is the benchmark dose.

Code	**Name**	**Extra risk Function**, $\boldsymbol{R_E(d)}$	**BMD**, $\boldsymbol{\xi}$
M1	Logistic	$\frac{1-\exp(-\beta_1 d)}{1+\exp(-\beta_0-\beta_1 d)}$	$\frac{\ln\left(\frac{1+BMRe^{-\beta_0}}{1-BMR}\right)}{\beta_1}$
M2	Probit	$\frac{\Phi(\beta_0+\beta_1 d)-\Phi(\beta_0)}{1-\Phi(\beta_0)}$	$\frac{\Phi^{-1}\{[1-\Phi(\beta_0)]BMR+\Phi(\beta_0)\}-\beta_0}{\beta_1}$
M3	Quantal-Linear	$1-\exp(-\beta_1 d)$	$\frac{-\ln(1-BMR)}{\beta_1}$
M4	Quantal-Quadratic	$1-\exp(-\beta_1 d^2)$	$\left(-\frac{\ln(1-BMR)}{\beta_1}\right)^{\frac{1}{2}}$
M5	Two-Stage	$1-\exp(-\beta_1 d-\beta_2 d^2)$	$\frac{-\beta_1+\sqrt{\beta_1^2-4\beta_2\ln(1-BMR)}}{2\beta_2}$
M6	Log-Logistic	$\frac{1}{1+\exp(-\beta_0-\beta_1\ln(d))}$	$\exp\left(\frac{\ln\left(\frac{BMR}{1-BMR}\right)-\beta_0}{\beta_1}\right)$
M7	Log-Probit	$\Phi(\beta_0+\beta_1\ln(d))$	$\exp\left(\frac{\Phi^{-1}(BMR)-\beta_0}{\beta_1}\right)$
M8	Weibull	$1-\exp(-e^{\beta_0}d^{\beta_1})$	$\exp\left(\frac{\ln(-\ln(1-BMR))-\beta_0}{\beta_1}\right)$

2.2 Reparameterization for quantal dose-response models

As mentioned in §1.5, in applications of Bayesian benchmark analysis to quantal data, objective and sometimes improper forms for the prior p.d.f. $\pi(\boldsymbol{\theta})$ are common, as indicated earlier. These typically appear as diffuse Gaussian priors on the β-parameters. From this, the joint posterior distribution for $\boldsymbol{\theta}$ is obtained using

Bayes formula. An advantage here is that objective priors are usually easy to apply: although they generally lead to intractable integrals, computer intensive operations such as Markov chain Monte Carlo (McMC) methods can produce a sample from the joint posterior of $\boldsymbol{\theta}$ (Robert and Casella, 2011). If the sample is sufficiently large and stable, the output can be used to approximate the posterior, from which inferences on the BMD may be conducted. As we suggest above, however, a disadvantage is that β-parameters may have unclear subject-matter interpretations if those parameters are not target quantities of interest. If informative prior information were available on the risk-analytic quantities under study, the ambiguous interpretation(s) of these traditional, regression-type parameterizations makes incorporation of such information more difficult. This may hinder effective application of the Bayesian approach in this benchmark setting.

For benchmark risk analysis, at least, it is plausible that substantive prior knowledge is available, but not in the form of information on a regression-type β-parameter. Instead, a risk assessor, toxicologist, or other domain expert would typically have prior knowledge about the target parameter, the BMD, and possibly also about other application-specific values such as the risk at certain doses. In order to conveniently derive and quantify this knowledge, parameterizations that are most familiar to the expert should be favored (Grieve, 1988).

The goal is to utilize the potential of the domain expert's prior knowledge for making inferences on the BMD. To do so, we reparameterize the risk function $R(d)$ in terms of meaningful parameters whose prior distributions are more intuitive to elicit in practice. The reparameterization strategy is not new, even in benchmark analysis; e.g., Parham and Portier (2005, §14.3.4) re-expressed the quantal-linear model in terms of the BMD to facilitate construction of BMDLs (under a frequentist framework). Following on their lead, we reformulate $\boldsymbol{\theta}$ in terms of well-understood risk-analytic quantities: for the dose-response models with two parameters in Table 2.1, they are reparameterized in terms of the target value, BMD (denoted in the

sequel as ξ) and the background risk, say, $\gamma_0 = R(0)$. Thus $\boldsymbol{\theta}$ becomes the vector $[\xi\ \gamma_0]^{\mathrm{T}}$.

For the models with three unknown parameters in Table 2.1, they are reparameterized with ξ, $\gamma_0 = R(0)$, and a parameter γ_1 defined as $R(d_\ell)$ for some non-zero dose level d_ℓ. Unless otherwise specified, d_ℓ is set to the highest dose, so $\gamma_1 = R(d_m)$. Thus, $\boldsymbol{\theta} = [\xi\ \gamma_0\ \gamma_1]^{\mathrm{T}}$. The latter two quantities are technically nuisance parameters as far as the BMD is concerned, but one or both are nonetheless likely to be associated with non-trivial prior information; e.g., historical control data may inform $\gamma_0 = R(0)$ (Wheeler and Bailer, 2012; Shao, 2012). The mathematical developments for the quantal-response models in Table 2.1 are as follows.

2.3 Reparameterizing the Logistic Model (M1)

The risk function for the logistic model from Table 2.1 is

$$R(d) = \frac{1}{1 + e^{-\beta_0 - \beta_1 d}}. \tag{2.3.1}$$

Clearly $\gamma_0 = R(0) = \frac{1}{1+e^{-\beta_0}}$. Our goal is to rewrite β_0 and β_1 in terms of γ_0 and ξ=BMD. Begin with γ_0. After some simple algebra we find

$$\beta_0 = \log\left(\frac{\gamma_0}{1-\gamma_0}\right). \tag{2.3.2}$$

Next, Wickens (2011) shows that the BMD for the logistic model at a fixed BMR $\in (0,1)$ is

$$\xi = \frac{\ln\left(\frac{1+\mathrm{BMR}e^{-\beta_0}}{1-\mathrm{BMR}}\right)}{\beta_1}. \tag{2.3.3}$$

If we solve Equation (2.3.3) for β_1, we find

$$\beta_1 = \frac{1}{\xi}\log\left(\frac{1+e^{-\beta_0}\mathrm{BMR}}{1-\mathrm{BMR}}\right). \tag{2.3.4}$$

Substitute (2.3.2) into (2.3.4) so that

$$\beta_1 = \frac{1}{\xi}\log\left(\frac{1+e^{-\text{logit}(\gamma_0)}\cdot \text{BMR}}{1-\text{BMR}}\right). \tag{2.3.5}$$

This defines β_0 and β_1 in terms of γ_0 and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into Equation (2.3.1):

$$\begin{aligned} R(d) &= \frac{1}{1+\exp\left[-\text{logit}(\gamma_0) - \frac{1}{\xi}\log\left(\frac{1+e^{-\text{logit}(\gamma_0)}\cdot \text{BMR}}{1-\text{BMR}}\right)\cdot d\right]} \\ &= \left\{1+\exp\left[-\text{logit}(\gamma_0) - \frac{1}{\xi}\log\left(\frac{1+e^{-\text{logit}(\gamma_0)}\cdot \text{BMR}}{1-\text{BMR}}\right)\cdot d\right]\right\}^{-1}. \end{aligned} \tag{2.3.6}$$

2.4 Reparameterizing the Probit Model (M2)

The risk function for the probit model from Table 2.1 is

$$R(d) = \Phi(\beta_0 + \beta_1 d). \tag{2.4.1}$$

Clearly $\gamma_0 = R(0) = \Phi(\beta_0)$. Our goal is to rewrite β_0 and β_1 in terms of γ_0 and ξ=BMD. Begin with γ_0. We have

$$\beta_0 = \Phi^{-1}(\gamma_0). \tag{2.4.2}$$

Next, Wickens (2011) shows that the BMD for the probit model at a fixed BMR $\in (0,1)$ is

$$\xi = \frac{\Phi^{-1}\{\text{BMR}[1-\Phi(\beta_0)] + \Phi(\beta_0)\} - \beta_0}{\beta_1}. \tag{2.4.3}$$

If we solve Equation (2.4.3) for β_1, we find

$$\beta_1 = \frac{\Phi^{-1}\{\text{BMR}[1-\Phi(\beta_0)] + \Phi(\beta_0)\} - \beta_0}{\xi}. \tag{2.4.4}$$

Substitute (2.4.2) into (2.4.4) so that

$$\beta_1 = \frac{\Phi^{-1}\{\text{BMR}[1-\gamma_0] + \gamma_0\} - \Phi^{-1}(\gamma_0)}{\xi}. \tag{2.4.5}$$

This defines β_0 and β_1 in terms of γ_0 and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into Equation (2.4.1):

$$\begin{aligned} R(d) &= \Phi(\beta_0 + \beta_1 d) \\ &= \Phi\left\{\Phi^{-1}(\gamma_0) + \frac{\Phi^{-1}\{\mathrm{BMR}[1-\gamma_0]+\gamma_0\} - \Phi^{-1}(\gamma_0)}{\xi} d\right\}. \end{aligned} \tag{2.4.6}$$

2.5 Reparameterizing the Quantal Linear Model (M3)

The risk function for the quantal linear model from Table 2.1 is

$$R(d) = 1 - \exp(-\beta_0 - \beta_1 d). \tag{2.5.1}$$

Clearly $\gamma_0 = R(0) = 1 - \exp(-\beta_0)$. Our goal is to rewrite β_0 and β_1 in terms of γ_0 and ξ=BMD. Begin with γ_0. We have

$$\beta_0 = -\log(1-\gamma_0). \tag{2.5.2}$$

Next, Wickens (2011) shows that the BMD for the quantal-linear model at a fixed BMR $\in (0,1)$ is

$$\xi = \frac{-\log(1-\mathrm{BMR})}{\beta_1}. \tag{2.5.3}$$

If we solve Equation (2.5.3) for β_1, we find

$$\beta_1 = \frac{-\log(1-\mathrm{BMR})}{\xi}. \tag{2.5.4}$$

This defines β_0 and β_1 in terms of γ_0 and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into Equation (2.5.1):

$$R(d) = 1 - \exp\left(\log(1-\gamma_0) + \frac{\log(1-\mathrm{BMR})}{\xi} d\right). \tag{2.5.5}$$

2.6 Reparameterizing the Quantal Quadratic Model (M4)

The risk function for the quantal quadratic model from Table 2.1 is

$$R(d) = \gamma_0 + (1 - \gamma_0)[1 - \exp(-\beta_1 d^2)]. \tag{2.6.1}$$

Clearly $\gamma_0 = R(0)$. Our goal is to rewrite β_1 in terms of γ_0 and ξ=BMD.

Next, Wickens (2011) shows that the BMD for the quantal-quadratic model at a fixed BMR $\in (0, 1)$ is

$$\xi = \left(-\frac{\log(1 - \text{BMR})}{\beta_1} \right)^{\frac{1}{2}}. \tag{2.6.2}$$

If we solve Equation (2.6.2) for β_1, we find

$$\beta_1 = -\frac{\log(1 - \text{BMR})}{\xi^2}. \tag{2.6.3}$$

Therefore, the corresponding, reparameterized risk function is obtained by substituting (2.6.3) into Equation (2.6.1):

$$\begin{aligned} R(d) &= \gamma_0 + (1 - \gamma_0)[1 - \exp(-\beta_1 d^2)] \\ &= \gamma_0 + (1 - \gamma_0) \left(1 - \exp\left\{ \frac{\log(1 - \text{BMR})}{\xi^2} d^2 \right\} \right). \end{aligned} \tag{2.6.4}$$

2.7 Reparameterizing the Multi-stage (Two-stage) Model (M5)

The risk function for the two-stage model from Table 2.1 is

$$R(d) = 1 - \exp(-\beta_0 - \beta_1 d - \beta_2 d^2). \tag{2.7.1}$$

With this, we define $\gamma_0 = R(0) = 1 - \exp(-\beta_0)$, hence, $\exp(-\beta_0) = 1 - \gamma_0$. Substitute $\exp(-\beta_0) = 1 - \gamma_0$ into the risk function to find

$$\begin{aligned} R(d) &= 1 - \exp(-\beta_0)\exp(-\beta_1 d - \beta_2 d^2) \\ &= 1 - (1-\gamma_0)\exp(-\beta_1 d - \beta_2 d^2) \\ &= 1 + (1-\gamma_0)(1 - 1 - \exp(-\beta_1 d - \beta_2 d^2)) \\ &= 1 + (1-\gamma_0)(1 - \exp(-\beta_1 d - \beta_2 d^2)) - (1-\gamma_0) \\ &= \gamma_0 + (1-\gamma_0)(1 - \exp(-\beta_1 d - \beta_2 d^2)). \end{aligned} \tag{2.7.2}$$

Next, we reexpress β_1 and β_2 in terms of ξ , γ_0 and γ_1. Begin with γ_1, expressed in terms of γ_0, β_1 and β_2. From the definition of an γ_1, we know

$$\gamma_1 = \gamma_0 + (1-\gamma_0)(1 - \exp(-\beta_1 d_m - \beta_2 d_m^2)),$$

where d_m denotes the m^{th} (the highest) dose level.

Denote $\Gamma_5 = \log\left(\frac{1-\gamma_1}{1-\gamma_0}\right)$. Recall that $0 < \gamma_1 < 1$. After some algebra, we obtain

$$\beta_2 d_m^2 + \beta_1 d_m + \Gamma_5 = 0. \tag{2.7.3}$$

Next, (Piegorsch and Bailer, 2005, Ex. 4.12) show that the benchmark dose for the two-stage model can be found to be

$$\xi = \frac{-\beta_1 + \sqrt{\beta_1^2 - 4\beta_2 \ln(1 - \text{BMR})}}{2\beta_2}. \tag{2.7.4}$$

Now, denote $C_5 = -\ln(1 - \text{BMR})$, from Equation (2.7.4), we have

$$\beta_2 \xi^2 + \beta_1 \xi - C_5 = 0. \tag{2.7.5}$$

If we solve Equation (2.7.3) for β_1, then we have $\beta_1 = \frac{-\Gamma_5 - \beta_2 d_m^2}{d_m}$. Substitute the result into (2.7.5) to find

$$\beta_2 = \frac{\Gamma_5 \xi + C_5 d_m}{\xi d_m (\xi - d_m)}. \tag{2.7.6}$$

Similarly, if we solve Equation (2.7.3) for β_2, then we have $\beta_2 = \frac{-\Gamma_5 - \beta_1 d_m}{d_m^2}$. Substitute the result into (2.7.5) to find

$$\beta_1 = \frac{C_5 d_m^2 + \Gamma_5 \xi^2}{\xi d_m (d_m - \xi)}. \tag{2.7.7}$$

So far we have rewritten β_1 and β_2 in terms of Γ_5 (or γ_0 and γ_1) and ξ . Therefore, the reparameterized risk function is obtained by substituting these expressions into Equation (2.7.2). This produces

$$\begin{aligned} R(d) &= \gamma_0 + (1-\gamma_0)\left(1 - \exp\left(-\frac{C_5 d_m^2 + \Gamma_5 \xi^2}{\xi d_m (d_m - \xi)} d - \frac{\Gamma_5 \xi + C_5 d_m}{\xi d_m (\xi - d_m)} d^2\right)\right) \\ &= \gamma_0 + (1-\gamma_0)\left(1 - \exp\left(\frac{C_5 d_m^2 d + \Gamma_5 \xi^2 d}{\xi d_m (\xi - d_m)} - \frac{\Gamma_5 \xi d^2 + C_5 d_m d^2}{\xi d_m (\xi - d_m)}\right)\right) \\ &= \gamma_0 + (1-\gamma_0)\left(1 - \exp\left(\frac{C_5 d_m d(d_m - d) + \Gamma_5 \xi d(\xi - d)}{\xi d_m (\xi - d_m)}\right)\right), \end{aligned} \tag{2.7.8}$$

where $\Gamma_5 = \log\left(\frac{1-\gamma_1}{1-\gamma_0}\right)$ and $C_5 = -\log(1 - \text{BMR})$.

2.8 Reparameterizing the Log-logistic Model (M6)

The risk function for the log-logistic model is

$$R(d) = \gamma_0 + (1-\gamma_0)\left[1 + \exp(-\beta_0 - \beta_1 \ln d)\right]^{-1}. \tag{2.8.1}$$

Clearly, $\lim_{d\to 0} R(d) = \gamma_0$. As above, our goal is to rewrite β_0 and β_1 in terms of γ_0, ξ and γ_1. Begin with γ_1, which is defined by the following equation:

$$\gamma_1 = \gamma_0 + (1-\gamma_0)\left[1 + \exp(-\beta_0 - \beta_1 \ln d_m)\right]^{-1},$$

where d_m denotes the m^{th} (the highest) dose level.

Denote $\Gamma_6 = \frac{1-\gamma_1}{1-\gamma_0}$, after some simple algebra we have

$$\beta_0 + \beta_1 \log d_m + \Gamma_6 = 0. \tag{2.8.2}$$

Next, Wickens (2011) shows that the BMD for the log-logistic model at a fixed BMR $\in (0,1)$ is

$$\xi = \exp\left(\frac{\text{logit}(\text{BMR}) - \beta_0}{\beta_1}\right). \tag{2.8.3}$$

Denote $C_6 = \log\left(\frac{\text{BMR}}{1-\text{BMR}}\right)$, after some simple algebra we have

$$\beta_0 + \beta_1 \log \xi - C_6 = 0. \tag{2.8.4}$$

If we solve Equation (2.8.2) for β_0, we find $\beta_0 = -\beta_1 \log d_m - \Gamma_6$, substitute the result into Equation (2.8.6) to find

$$\beta_1 = \frac{C_6 + \Gamma_6}{\log \xi - \log d_m}. \tag{2.8.5}$$

Similarly, if we solve Equation (2.8.2) for β_1, we find $\beta_1 = \frac{-\beta_0 - \Gamma_6}{\log d_m}$, substitute the result into Equation (2.8.6) to find

$$\beta_0 = \frac{C_6 \log d_m + \Gamma_6 \log \xi}{\log d_m - \log \xi}. \tag{2.8.6}$$

This defines β_0 and β_1 in terms of Γ_6 (or γ_0 and γ_1) and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into Equation (2.8.1):

$$\begin{aligned} R(d) &= \gamma_0 + (1-\gamma_0)\left[1 + \exp\left(-\frac{C_6 \log d_m + \Gamma_6 \log \xi}{\log d_m - \log \xi} - \frac{C_6 + \Gamma_6}{\log \xi - \log d_m} \log d\right)\right]^{-1} \\ &= \gamma_0 + (1-\gamma_0)\left[1 + \exp\left(\frac{C_6(\log d_m - \log d) + \Gamma_6(\log \xi - \log d)}{\log \xi - \log d_m}\right)\right]^{-1} \end{aligned} \tag{2.8.7}$$

where $\Gamma_6 = \log\left(\frac{1-\gamma_1}{1-\gamma_0}\right)$ and $C_6 = \log\left(\frac{\text{BMR}}{1-\text{BMR}}\right)$.

2.9 Reparameterizing the Log-probit Model (M7)

The risk function for the log-probit model is

$$R(d) = \gamma_0 + (1-\gamma_0)\Phi(\beta_0 + \beta_1 \ln d). \tag{2.9.1}$$

Clearly, $\lim_{d\to 0} R(d) = \gamma_0$. As above, our goal is to rewrite β_0 and β_1 in terms of γ_0, ξ and γ_1. Begin with γ_1, which is defined by the following equation:

$$\gamma_1 = \gamma_0 + (1 - \gamma_0)\Phi(\beta_0 + \beta_1 \ln d_m),$$

where d_m denotes the m^{th} (the highest) dose level.

Denote $\Gamma_7 = \Phi^{-1}\left(\frac{\gamma_1-\gamma_0}{1-\gamma_0}\right)$, after some simple algebra we have

$$\beta_0 + \beta_1 \log d_m - \Gamma_7 = 0. \tag{2.9.2}$$

Next, derivations similar to those in model M6 show that the BMD for the log-probit model at a fixed BMR $\in (0, 1)$ takes the same general form:

$$\xi = \exp\left(\frac{C_7 - \beta_0}{\beta_1}\right), \tag{2.9.3}$$

Denote $C_7 = \Phi^{-1}(\text{BMR})$, after some simple algebra we have

$$\beta_0 + \beta_1 \log \xi - C_7 = 0. \tag{2.9.4}$$

If we solve Equation (2.9.2) for β_0, we find $\beta_0 = \Gamma_7 - \beta_1 \log d_m$, substitute the result into Equation (2.9.4) to find

$$\beta_1 = \frac{C_7 - \Gamma_7}{\log \xi - \log d_m}. \tag{2.9.5}$$

Similarly if we solve Equation (2.9.2) for β_1, we find $\beta_1 = \frac{\Gamma_7-\beta_0}{\log d_m}$, substitute the result into Equation (2.9.4) to find

$$\beta_0 = \frac{C_7 \log d_m - \Gamma_7 \log \xi}{\log d_m - \ln \xi}. \tag{2.9.6}$$

This defines β_0 and β_1 in terms of Γ_7 (or γ_0 and γ_1) and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into

Equation (2.9.1):

$$\begin{aligned} R(d) &= \gamma_0 + (1-\gamma_0)\Phi\left(\frac{C_7 \log d_m - \Gamma_7 \log \xi}{\log d_m - \ln \xi} + \frac{C_7 - \Gamma_7}{\log \xi - \log d_m} \log d\right) \\ &= \gamma_0 + (1-\gamma_0)\Phi\left(\frac{C_7 \log d_m - \Gamma_7 \log \xi + \Gamma_7 \log d - C_7 \log d}{\log d_m - \log \xi}\right) \\ &= \gamma_0 + (1-\gamma_0)\Phi\left(\frac{C_7(\log d_m - \log d) + \Gamma_7(\log d - \log \xi)}{\log d_m - \log \xi}\right), \end{aligned} \tag{2.9.7}$$

where $\Gamma_7 = \Phi^{-1}\left(\frac{\gamma_1-\gamma_0}{1-\gamma_0}\right)$ and $C_7 = \Phi^{-1}(\text{BMR})$.

2.10 Reparameterizing the Weibull Model (M8)

The risk function for the Weibull model is

$$R(d) = \gamma_0 + (1-\gamma_0)\left[1 - \exp(-\exp(\beta_0 + \beta_1 \ln d))\right]. \tag{2.10.1}$$

Clearly, $\lim_{d\to 0} R(d) = \gamma_0$. As above, our goal is to rewrite β_0 and β_1 in terms of γ_0, ξ and γ_1. Begin with γ_1, which is defined by the following equation:

$$\gamma_1 = \gamma_0 + (1-\gamma_0)\left[1 - \exp(-\exp(\beta_0 + \beta_1 \ln d_m))\right],$$

where d_m denotes the m^{th} (the highest) dose level.

Denote $\Gamma_8 = \log\left(-\log\left(\frac{1-\gamma_1}{1-\gamma_0}\right)\right)$, after some simple algebra we have

$$\beta_0 + \beta_1 \log d_m - \Gamma_8 = 0. \tag{2.10.2}$$

Next, derivations similar to those in model M6 show that the BMD for the log-probit model at a fixed BMR $\in (0,1)$ takes the same general form:

$$\xi = \exp\left(\frac{C_8 - \beta_0}{\beta_1}\right), \tag{2.10.3}$$

Denote $C_8 = \log(-\log(1-\text{BMR}))$, after some simple algebra we have

$$\beta_0 + \beta_1 \log \xi - C_8 = 0. \tag{2.10.4}$$

If we solve Equation (2.10.2) for β_0, we find $\beta_0 = \Gamma_8 - \beta_1 \log d_m$, substitute the result into Equation (2.10.4) to find

$$\beta_1 = \frac{C_8 - \Gamma_8}{\log \xi - \log d_m}. \tag{2.10.5}$$

Similarly if we solve Equation (2.10.2) for β_1, we find $\beta_1 = \frac{\Gamma_8 - \beta_0}{\log d_m}$, substitute the result into Equation (2.10.4) to find

$$\beta_0 = \frac{C_8 \log d_m - \Gamma_8 \log \xi}{\log d_m - \ln \xi}. \tag{2.10.6}$$

This defines β_0 and β_1 in terms of Γ_8 (or γ_0 and γ_1) and ξ. The corresponding, reparameterized risk function is obtained by substituting these expressions into Equation (2.10.1):

$$\begin{aligned} & R(d) \\ = \;& \gamma_0 + (1-\gamma_0)\left[1 - \exp\left(-\exp\left(\frac{C_8 \log d_m - \Gamma_8 \log \xi}{\log d_m - \ln \xi} + \frac{C_8 - \Gamma_8}{\log \xi - \log d_m} \log d\right)\right)\right] \\ = \;& \gamma_0 + (1-\gamma_0)\left[1 - \exp\left(-\exp\left(\frac{C_8 \log d_m - \Gamma_8 \log \xi + \Gamma_8 \log d - C_8 \log d}{\log d_m - \log \xi}\right)\right)\right] \\ = \;& \gamma_0 + (1-\gamma_0)\left[1 - \exp\left(-\exp\left(\frac{C_8(\log d_m - \log d) + \Gamma_8(\log d - \log \xi)}{\log d_m - \log \xi}\right)\right)\right], \end{aligned} \tag{2.10.7}$$

where $\Gamma_8 = \log\left(-\log\left(\frac{1-\gamma_1}{1-\gamma_0}\right)\right)$ and $C_8 = \log(-\log(1 - \text{BMR}))$.

These various reparameterized forms explicitly display the eight dose-response functions from Table 2.1 in terms of model parameters pertinent to a toxicologist or risk assessor. These reparameterizations present more burdensome notation for $R(d)$. The explicit incorporation of the target parameter ξ and well-understood quantities such as γ_0 and γ_1 allows us, however, to formulate a more application-oriented hierarchical model, from which to produce inferences on ξ.

Chapter 3

Hierarchical Bayesian Benchmark Analysis

3.1 Prior Specification

Under our reformulation for the quantal-response models in Table 2.1, the unknown parameter vector becomes $\boldsymbol{\theta} = [\xi \ \gamma_0]^T$ for the two-parameter models and $\boldsymbol{\theta} = [\xi \ \gamma_0 \ \gamma_1]^T$ for the three-parameter models for use in the joint p.m.f. $f(\boldsymbol{Y}|\boldsymbol{\theta})$. To construct a Bayesian hierarchy, a joint p.d.f. to $\boldsymbol{\theta}$: $\pi(\boldsymbol{\theta}) = \pi(\xi, \gamma_0)$ is assigned for the two-parameter models and $\pi(\boldsymbol{\theta}) = \pi(\xi, \gamma_0, \gamma_1)$ for the three-parameter models. Mimicking previous Bayesian models for benchmark analysis (Shao and Small, 2011, 2012; Shao, 2012), the unknown parameters are assumed to enter into the prior independently, so that $\pi(\xi, \gamma_0) = \pi(\xi)\pi(\gamma_0)$ or $\pi(\xi, \gamma_0, \gamma_1) = \pi(\xi)\pi(\gamma_0)\pi(\gamma_1)$.

To specify the individual components of the joint prior, established yet flexible forms are employed. For the non-negative quantity ξ, we model its distribution using an inverse gamma: $\xi \sim IG(\alpha, \beta)$ with marginal prior density $\pi(\xi|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\xi^{-(\alpha+1)}e^{-\beta/\xi}I_{(0,\infty)}(\xi)$, where $\Gamma(a)$ is the usual gamma function and $I_{\mathbb{A}}(x)$ is the indicator function that returns 1 if $x \in \mathbb{A}$ and 0 otherwise. For the probability parameter $\gamma_0 = R(0)$, $\gamma_0 \sim Beta(\psi, \omega)$ is taken with marginal prior $\pi(\gamma_0|\psi, \omega) = \frac{\Gamma(\psi+\omega)}{\Gamma(\psi)\Gamma(\omega)}\gamma_0^{\psi-1}(1-\gamma_0)^{\omega-1}I_{(0,1)}(\gamma_0)$. Where needed (models M5–M8), $\gamma_1 \sim Beta(\kappa, \lambda)$ is similarly set.

3.2 Prior Elicitation

The various hyperparameters, α, β, ψ, ω, κ and λ, control the nature and form of the prior densities, and these require complete specification for implementation of the model as it is proposed. (One could also build further levels into the hierarchy by constructing hyperprior p.d.f.s, and this is an area of potential future investigation.

However, this is not studied here.) The goal is then to fully *elicit* each marginal prior by incorporating the domain expert's prior knowledge of the associated quantities. A broad literature exists on how to conduct prior elicitation, see, e.g., O'Hagan *et al.* (2006) or Kuhnert (2011) and the references therein. This is not summarized here, however, the general strategy is taken from it that domain experts with minimal statistical expertise are best able to provide prior information in the form of basic location summaries, e.g., population means, medians, quartiles, percentiles, etc. Quartiles can be especially effective: medians, i.e., 2nd quartiles, are often recovered effectually from domain experts when eliciting measures of central tendency. (With right-skewed quantities such as the BMD, observers tend to underestimate population means.) Thus whenever possible a median specification is always included when eliciting features of the prior p.d.f.s. Since in the hierarchical model each prior contains two hyper-parameters, a second, separate prior quartile specification is also required to complete the elicitation. Based on interactions with toxicologists and risk assessment domain experts, for the target parameter ξ, specification of the first/lower quartile (Q_1) along with the median (Q_2) of the IG prior is found to be easiest to achieve. Put simply, prior expert knowledge for ξ is likely to be more accurate closer to the origin, since BMDs are associated with adverse effects at low doses. Similarly, for the Beta prior on γ_0 and γ_1 the two quartiles Q_1 and Q_2 are also elicited. After that, the pertinent hyper-parameters given the two quartiles are then solved. Of course, other elicitation values are possible with these sorts of prior densities, and analysts may wish to experiment with selection of other quartiles, terciles, percentiles, etc. Take the two quantities, Q_1 and Q_2, for example. The technical aspects on derivation of the prior parameters α, β, ψ, ω, κ and λ are given here for $\xi \sim IG(\alpha, \beta)$, $\gamma_0 \sim Beta(\psi, \omega)$ and $\gamma_1 \sim Beta(\kappa, \lambda)$ with elicited quartiles $Q_{1\xi}, Q_{2\xi}$ for ξ, and $Q_{1\gamma_0}, Q_{2\gamma_0}$ for γ_0 and $Q_{1\gamma_1}, Q_{2\gamma_1}$ for γ_1. Start with the elicitation for ξ: by definition, $Q_{1\xi}$ and $Q_{2\xi}$ satisfy

$$\int_0^{Q_{j\xi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \xi^{(-\alpha-1)} \exp\left\{-\frac{\beta}{\xi}\right\} d\xi = \frac{j}{4},$$

$(j = 1, 2)$, where the integrand is the IG p.d.f. This establishes a system of non-linear equations for α and β. Unfortunately, no closed-form solution exists for the system, and so numerical method is applied. In particular, a gradient method discussed by Barzilai and Borwein (1988) is employed and implemented in the `R` statistical environment (R Development Core Team, 2012) via the `BB` package (Varadhan and Gilbert, 2009). Simply put, this method uses iterative updating until half the L_2 norm of the system at the proposed solution is smaller than 10^{-10}. This root-finder requires initial guesses for the solutions. For α and β, initial values are both chosen to 0.001 (as described later, the hyperparameters for objective IG prior); for ψ, ω, κ and λ, they are set to 0.5 (again, the hyperparameters for an objective Beta prior). In most case, iterative solutions are converged within 100 steps with these initial guesses. Whenever convergence fails, the initial guesses will be updated with the results obtained from the last attempt and loaded to initialize a new search. A total of 10 attempts is allowed, however experience shows roots can be found within 2 attempts in most cases.

Similarly, for γ_0 the elicited quartiles $Q_{1\gamma_0}$ and $Q_{2\gamma_0}$ satisfy

$$\int_0^{Q_{j\gamma_0}} \frac{\Gamma(\psi+\omega)}{\Gamma(\psi)\Gamma(\omega)} \gamma_0^{\psi-1}(1-\gamma_0)^{\omega-1} d\gamma_0 = \frac{j}{4},$$

$(j = 1, 2)$, now applying the p.d.f. for the Beta prior. Here again, the resulting system of non-linear equations possesses no closed-form solution, so the `BB` package in `R` is again used. The hyperparameters for elicited $\pi(\gamma_1)$ are found in the same process.

3.3 Objective Priors

Cases can arise where the prior elicitation breaks down, say, if prior experience with a suspect toxin, its sister compounds, or its metabolites is so limited that the toxicologist simply has no idea where the BMD will lie. When this occurs, moving to objective specifications for the prior densities may be necessary. Many possibilities

exist for building objective priors on a strictly positive quantity such as ξ, and on a probability such as γ_0. For example, one could appeal to a joint bivariate Jeffreys prior and take $\pi(\boldsymbol{\theta}) \propto \sqrt{\det(E[\{\partial \log L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}^2])}$ (Jeffreys, 1961). The multi-parameter construction here is highly complex, however, and can lead to unstable posterior inferences in settings where one of the quantities represents a secondary, nuisance variable (Bernardo, 1979). Instead, a simpler approach is used.

An objective prior on ξ has been chosen to $\xi \sim IG(0.001, 0.001)$ [other options could be $\xi \sim Gamma(0.001, 0.001)$ as studied below]. The IG prior is a popular suggestion in the literature for right-skewed, positive quantities (Lambert *et al.*, 2005; Christensen *et al.*, 2011, §1.2), such as the BMD. One could alternatively appeal to an (improper) objective prior of the form $\pi(\xi) \propto 1/\xi$, a conventional choice that is often recommended for positive parameters (O'Hagan, 1994, §9.17). The improper prior assigns decaying density to values increasing from 0. This is reasonable since the low-dose extrapolated quantity BMD is more likely to be a small quantity. In order to facilitate the posterior computations, however, employing proper prior densities is preferred. In fact, experiences show the $IG(0.001, 0.001)$ prior can approximate the improper reciprocal prior quite well when ξ is not too close to 0. The $Gamma(0.001, 0.001)$ can also serve as an approximator for $1/\xi$, although for large dose values the approximation can break down. Indeed, it is found that scaling the doses to make the highest dose equal 1 is a convenient device to allow both $IG(0.001, 0.001)$ and $Gamma(0.001, 0.001)$ to approximate the conventional improper $1/\xi$ prior. As such, hierarchical calculations are performed with doses scaled so that the maximum dose equals 1.

For an objective prior on γ_0 and γ_1, one could apply the simple Bayes-Laplace uniform priors, $unif(0, 1)$. However, γ_0 quantifies the probability of adverse response under no exposure, and is typically not very large in risk-analytic applications; while γ_1 quantifies the probability of adverse repose under highest exposure, and is typically not very small. It therefore seems unrealistic to assign them uniform prior densities

across all values in the unit interval, especially those approaching 1 for γ_0, and 0 for γ_1.

An alternative choice would be the univariate Jeffreys prior: $Beta\left(\frac{1}{2}, \frac{1}{2}\right)$. This is a conventional, objective prior for proportions, which assigns symmetrically high density to values close to 0 and values close to 1. The former is reasonable for γ_0 in our toxicological setting, although the latter is problematic, since γ_0 represents the probability of an adverse response under no exposure. This value is not typically large in risk-analytic applications. Similarly, for γ_1, assigning high density to values close to 0 is problematic. In the end, even the valid-but-imprecise argument that low prior weight should be assigned to values of $\gamma_0 \to 1$ and values of $\gamma_1 \to 0$ is itself a form of subjective prior knowledge, and construction of truly objective priors for γ_0 and γ_1 may be difficult. However, both γ_0 and γ_1 (or, $R(d_\ell)$, risk at dose level, ℓ) are well-understood quantity in practice, and as others have noted there will often be at least some useful historical information available for γ_0 and γ_1. It is expected that calls for objective priors on γ_0 and γ_1 in this setting will be rare. If in the extreme this is not the case, default is made to the objective Jeffreys priors for γ_0 and γ_1 $\sim Beta\left(\frac{1}{2}, \frac{1}{2}\right)$, with the recognition of the enigmatic aspects of such a strategy.

3.4 Posterior Analysis via Stochastic Approximation

Given the priors specified in §3.1, and employing the joint binomial p.m.f. (1.3.1), the joint posterior p.d.f. for $\boldsymbol{\theta} = [\xi \ \gamma_0]^T$ is

$$\begin{aligned} & \pi(\xi, \gamma_0 | \boldsymbol{Y}) \\ = \ & \frac{f(\boldsymbol{Y}|\xi, \gamma_0)\pi(\xi|\alpha, \beta)\pi(\gamma_0|\psi, \omega)}{m(\boldsymbol{Y})} \\ = \ & \frac{\prod_{i=1}^{n} \binom{N_i}{Y_i} R(d_i)^{Y_i}[1 - R(d_i)]^{N_i - Y_i}}{m(\boldsymbol{Y})} \frac{\beta^\alpha e^{-\beta/\xi}}{\Gamma(\alpha)\xi^{\alpha+1}} \frac{\Gamma(\psi+\omega)}{\Gamma(\psi)\Gamma(\omega)} \gamma_0^{\psi-1}(1-\gamma_0)^{\omega-1}. \end{aligned} \tag{3.4.1}$$

And, the joint posterior p.d.f for $\boldsymbol{\theta} = [\xi \ \gamma_0 \ \gamma_1]^T$ is

$$\begin{aligned} & \pi(\xi, \gamma_0, \gamma_1 | \boldsymbol{Y}) \\ = & \frac{f(\boldsymbol{Y}|\xi, \gamma_0, \gamma_1)\pi(\xi|\alpha, \beta)\pi(\gamma_0|\psi, \omega)\pi(\gamma_1|\kappa, \lambda)}{m(\boldsymbol{Y})} \\ = & \frac{\prod_{i=1}^{n} \binom{N_i}{Y_i} R(d_i)^{Y_i}[1 - R(d_i)]^{N_i - Y_i}}{m(\boldsymbol{Y})} \frac{\beta^\alpha e^{-\beta/\xi}}{\Gamma(\alpha)\xi^{\alpha+1}} \frac{\Gamma(\psi+\omega)}{\Gamma(\psi)\Gamma(\omega)} \gamma_0^{\psi-1}(1-\gamma_0)^{\omega-1} \\ & \frac{\Gamma(\kappa+\lambda)}{\Gamma(\lambda)\Gamma(\kappa)} \gamma_1^{\kappa-1}(1-\gamma_1)^{\lambda-1}. \end{aligned} \tag{3.4.2}$$

As mentioned in §1.4, estimation and inferences about $\boldsymbol{\theta}$ follow directly from its posterior distribution. These usually require several steps of integrations to, for example, 'marginalize' the joint posterior p.d.f, find certain moments of the marginal posterior p.d.f (e.g., the posterior mean), or find certain percentile of the marginal posterior p.d.f., etc. Besides these, some unknown constants in the joint posterior p.d.f may also require integration. For example, the denominators of (3.4.1) and (3.4.2) contain the marginal likelihoods

$$m(\boldsymbol{Y}) = \int_0^\infty \int_0^1 f(\boldsymbol{Y}|\xi, \gamma_0)\pi(\xi|\alpha, \beta)\pi(\gamma_0|\psi, \omega) d\gamma_0 d\xi,$$

and

$$m(\boldsymbol{Y}) = \int_0^\infty \int_0^1 \int_0^1 f(\boldsymbol{Y}|\xi, \gamma_0, \gamma_1)\pi(\xi|\alpha, \beta)\pi(\gamma_0|\psi, \omega)\pi(\gamma_1|\kappa, \lambda) d\gamma_0 \gamma_1 d\xi.$$

which are integrals as well.

These integrals greatly hinder the estimation/inference process. For instance, in order to find a marginal posterior median for ξ in $\pi(\xi|\boldsymbol{Y})$ from a three-parameter model. First, $m(\boldsymbol{Y})$ needs to be evaluated through a double integral. Second, $\pi(\xi|\boldsymbol{Y})$ is obtained by 'integrating out' the γ_0 and γ_1 components from $\pi(\xi, \gamma_0, \gamma_1|\boldsymbol{Y})$ (another double integral). Third, the median t is obtained by solving integration equation $\int_0^t \pi(\xi|\boldsymbol{Y})d\xi = \frac{1}{2}$. To directly evaluate integrals, Zellner and Rossi (1984) summarize four methods, namely: (1) analytical integration, (2) asymptotic expansion, (3) numerical integration and (4) Monte Carlo integration. However, experience shows that

for posterior p.d.f.s such as (3.4.1) and (3.4.2), analytical integration is to difficult and asymptotic expansions are too complicated for convenient use. Because of these issues, focus here is switched to computationally intensive posterior approximations.

Among many of those, Markov chain Monte Carlo (McMC) approaches are perhaps the most popular ones nowadays and have been applied with great success to problems such as this (Barnes, 2012; Robert and Casella, 2011). By using proper priors in (3.4.1) and (3.4.2), the posterior densities are guaranteed to be integrable (Gelman *et al.*, 2004, §5.3), and appropriate McMC simulations for them become possible. (This motivates the preference for the IG prior for ξ and Beta priors for γ_0 and γ_1, respectively.) Unfortunately, the complexity of the posteriors still prevents using some well-known McMC algorithms such as Gibbs sampling (Gelfand, 2000). The conditional posterior distributions [e.g., $\pi(\gamma_0|\xi, \boldsymbol{Y})$ in two-parameter models and $\pi(\xi|\gamma_0, \gamma_1, \boldsymbol{Y})$ in three-parameter models] are all not in common distribution, sampling from them are difficult and some popular algorithm such as the adaptive rejection sampling method in Gibbs sampler (Gilks and Wild, 1992) still fail to sample them properly. Furthermore, experiences also show that the hybrid Metropolis within Gibbs sampler algorithm (Robert and Casella, 2004, §10.3) also cannot provide desirable results. Facing these difficulties, the more general Metropolis-Hastings approaches (Robert and Casella, 2004, Ch. 7) are thus considered. A difficulty for this type of algorithms, however, is to find proper tuning parameters for the variance-covariance matrix of the multivariate proposal density. Sometimes, maximum likelihood estimates for the variance-covariance matrix provide some insights. (see `R`, `MCMCPack` package for example.) However, MLE can be unstable for some quantal-response data, therefore, a simpler and less MLE-dependant method is preferred. Because of this, *adaptive* Metropolis (AM) algorithms are considered to automate the process of looking for proper tuning parameters. These tune the variance of the underlying Metropolis proposal density adaptively when generating ongoing draws of the chain. From the experience with a variety of such methods [for representative examples, see

Haario *et al.* (2001); Roberts and Rosenthal (2009)], a global AM procedure with componentwise adaptive scaling described by Andrieu and Thoms (2008) is favored. This is introduced in detail in the following section.

3.5 Global Adaptive Metropolis Algorithm With Componentwise Adaptive Scaling

The 'AM6' algorithm in Andrieu and Thoms (2008) provides a global adaptive metropolis algorithm with componentwise adaptive scaling. It uses all previous iterations in the chain to update the variance-covariance matrix of the current multivariate proposal density. The operation is global in the sense that all the parameters (here, ξ and γ_0 for the two-parameter models and ξ, γ_0 and γ_1 for the three-parameter models) are updated simultaneously at each bivariate or trivariate draw. It is componentwise in the sense that the updating step sizes for ξ and γ_0 (or ξ, γ_0 and γ_1 for the three-parameter models) are controlled by separate adaptive scaling parameters. Other mathematical backgrounds for this algorithm, such as a discussion on ergodicity and convergence issues, are illustrated in Andrieu and Thoms (2008).

For practical purposes, however, the acceptability of the data is first checked before employing the algorithm. Experiences show that a shallow dose response can create unstable frequentist $\hat{\xi}_{100\text{BMR}}$ estimates, and very shallow responses may cause the iterative model fit to fail entirely. Indeed, the EPA's BMDS software program will not estimate a BMD for a flat or negatively trending dose response (Wheeler and Bailer, 2009a). To check the acceptability of the data before employing the algorithm, a simple screen is performed as follows: Over increasing doses $0 = d_1 < \ldots < d_i < \ldots < d_m$ $(i = 1, \ldots m)$, the empirical extra risks are first calculated:

$$\tilde{R}_E(d_i) = \frac{\frac{Y_i}{N_i} - \frac{Y_1}{N_1}}{1 - \frac{Y_1}{N_1}}.$$

After that, each point $(d_i, \tilde{R}_E(d_i))$ is connected to the origin point and the largest slope, $S_{\max}$, is found among all $m - 1$ rays. If $S_{\max} \leq 0$, no increasing trend is

evidenced and the corresponding data set is marked as a 'data failure'. Notice that, no formal trend test to detect a decreasing trend (Wheeler and Bailer, 2009a) is performed here, but what is required is that at least one estimated risk at some d_i ($i > 1$) is higher than the estimated background risk.

For data passing the simple 'data failure' screen as mentioned above, posterior simulation is performed via the 'AM6' algorithm in Andrieu and Thoms (2008). Denote $\boldsymbol{\theta}$ as the parameter vector ($[\xi\ \gamma_0]^{\mathrm{T}}$ for the two-parameter case or $[\xi\ \gamma_0\ \gamma_1]^{\mathrm{T}}$ for the three-parameter case). Let $v_{k,1}, \ldots, v_{k,U}$ be adaptive scaling parameters and set $V_k = \mathrm{diag}\,(v_{k,1}, \ldots, v_{k,U})$ at the kth draw in the Monte Carlo chain ($k = 1, \ldots, K$). (U denotes the dimension for the parameter vector; $U = 2$ for our two-parameter models and $U = 3$ for the three-parameter models.) Take α as the acceptance probability in the normal random-walk Metropolis sampler, i.e.,

$$\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}_k + Z_{k+1}) = \min\left(1, \frac{\pi(\boldsymbol{\theta}_k|\boldsymbol{Y})}{\pi(\boldsymbol{\theta}_k + Z_{k+1}|\boldsymbol{Y})}\right).$$

For notation in what follows, take e_u as a length-U canonical vector with the uth element equal to 1 and all others 0. Denote s_k as a step-size parameter that makes $\{s_k\}_{k=1}^{\infty}$ a decaying sequence; for example, the deterministic specification $s_k = 1/k^p$, with $p > 0$ is employed here. [Andrieu and Thoms (2008) recommend that $p > 1/2$, while Vihola (2012) suggests $p = 2/3$. Here, Vihola's suggestion is followed and $s_k = k^{-2/3}$ is used.] A brief description of the AM algorithm follows.

- Find starting values for ξ, γ_0, and (where needed) γ_1 in the chain as follows: take the starting point for γ_0 as $\frac{Y_1+0.25}{N_1+0.5}$. Similarly, take the starting point for γ_1 as $\frac{Y_j+0.25}{N_j+0.5}$, where index j indicates the j^{th} data observation at which $S_{\max}$ is attained. For the two-parameter models, the starting point for ξ is taken as $\mathrm{BMR}/S_{\max}$. This is obtained by first treating $R_E(d) = S_{\max}d$ as the approximate extra risk function and then solving for d in the equation $\mathrm{BMR} = S_{\max}d$. For the three-parameter models, the starting point for ξ is

taken as BMR/ $\left(\frac{\gamma_{10}-\gamma_{00}}{1-\gamma_{00}}\right)$ where γ_{00} and γ_{10} are the starting values for γ_0 and γ_1, respectively. These starting points collectively become the starting vector for the chain. Denote this as $\boldsymbol{\theta}_1$ in the algorithm and let $\mu_1 = \boldsymbol{\theta}_1$.

- Following Andrieu and Thoms (2008), let $\Sigma_1 = \mathbf{I}$ where $\mathbf{I}$ is the $U \times U$ identity matrix. And then, let $v_{1,1} = \ldots = v_{1,U} = \frac{2.38^2}{U}$ be the initial scaling parameters. Let $V_1 = \text{diag}\,(v_{1,1}, \ldots, v_{1,U})$, then $V_1^{1/2}\mathbf{I}V_1^{1/2}$ is the initial variance-covariance matrix, for use in the joint proposal distribution.

- Repeat for $k = 2, \ldots, K$:

  1. Given μ_{k-1}, Σ_{k-1} and $v_{k-1,1}, \ldots, v_{k-1,U}$, sample from a multivariate normal proposal distribution: $Z_k \sim N(0, V_{k-1}^{1/2}\Sigma_{k-1}V_{k-1}^{1/2})$ and set $\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + Z_k$ with probability $\alpha(\boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k-1} + Z_k)$, otherwise set $\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1}$.

  2. Update as follows for $u = 1, \ldots, U$:

$$\log(v_{k,u}) = \log(v_{k-1,u}) + s_k[\alpha(\boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k-1} + Z_k(u)e_u) - \bar{\alpha}_{**}],$$

     where $\bar{\alpha}_{**}$ is the acceptance rate of the candidate points generated at each iteration. Andrieu and Thoms (2008) suggest $\bar{\alpha}_{**} = 0.44$ provides good performance; that choice is made here.

  3. Update as follows for μ_k and Σ_k:

$$\mu_k = \mu_{k-1} + s_k(\boldsymbol{\theta}_k - \mu_{k-1}),$$

$$\Sigma_k = \Sigma_{k-1} + s_k[(\boldsymbol{\theta}_k - \mu_{k-1})(\boldsymbol{\theta}_k - \mu_{k-1})^T - \Sigma_{k-1}]$$

3.6 Monte Carlo 'burn-in' diagnostics

To account for potential instability in the early portions of the bivariate/trivariate AM chain of length K from §3.5, a 'burn-in' for the Monte Carlo sample is included (Gelman *et al.*, 2004, §11.6). This is coupled with the larger question of how to assess

the chain's convergence. Many approaches exist for diagnosing MCMC convergence (Cowles and Carlin, 1996), and indeed, the issue is a topic of ongoing debate (Robert and Casella, 2004, §12.2). Here a method mimicking Geweke (1992) is applied, where early portions of the chain are sub-sampled and compared against latter portions to determine where the larger sample of K draws begins to exhibit stable performance. To approximate independence between the two sub-samples, the chain is bifurcated by a gap of no less than $\frac{K}{5}$ consecutive draws.

For summary diagnostics, the difference in arithmetic means between the two bifurcated sub-samples is calculated, and each difference is divided by its standard error to produce a Z-statistic. The approximate standard error is taken as the square root of the sum of the variances of each mean. This is done separately for the ξ_k, γ_{0k}, and γ_{1k} components. To adjust for possible autocorrelation within each sub-sample, the individual variances are based on estimated spectral densities at frequency zero. Two parametric methods of estimating these spectral densities are considered here. One method is discussed by Heidelberger and Welch (1981), which consists of fitting a generalized linear model to the periodogram of the batched time series. This is implemented using `R`, via the `coda` package's `spectrum0` function. The second method first fits the periodogram with an autogressive model, and then use the spectral density of the fitted AR model to approximate the spectral density at frequency zero. Again, this method is implemented using `R`, now with the `coda` package's `spectrum0.ar` function. These two methods produce similar estimates of the spectral density at zero. The Heidelberger and Welch method is fast and is the default method in the `coda` package, hence is set as default here. The fit may fail due to high autocorrelation of the chain, however, and when this happens the AR method is switched to for estimating the spectral density at zero.

To monitor if the pattern of association between any two parameters within the larger chain also exhibits reasonable stability, comparison of up to three pairs of covariances between two sub-samples is included (for the two-parameter model, only

covariance between ξ and γ_0 is considered). As a diagnostic measure here, the sample covariances from each sub-sample are used

$$\widehat{\mathrm{Cov}}(\xi, \gamma_0) = \frac{1}{L}\sum_{k=1}^{L}(\xi_k - \bar{\xi})(\gamma_{0k} - \bar{\gamma}_0),$$

$$\widehat{\mathrm{Cov}}(\xi, \gamma_1) = \frac{1}{L}\sum_{k=1}^{L}(\xi_k - \bar{\xi})(\gamma_{1k} - \bar{\gamma}_1),$$

$$\widehat{\mathrm{Cov}}(\gamma_0, \gamma_1) = \frac{1}{L}\sum_{k=1}^{L}(\gamma_{0k} - \bar{\gamma}_0)(\gamma_{1k} - \bar{\gamma}_1),$$

where $\bar{\xi}$, $\bar{\gamma}_0$ and $\bar{\gamma}_1$ are the pertinent arithmetic means within a bifurcated sub-sample of length L (see below). To find the approximate variance of each estimated covariance, the following quantities are calculated

$$\tau_k(\xi, \gamma_0) = (\xi_k - \bar{\xi})(\gamma_{0k} - \bar{\gamma}_0),$$

$$\tau_k(\xi, \gamma_1) = (\xi_k - \bar{\xi})(\gamma_{0k} - \bar{\gamma}_1),$$

$$\tau_k(\gamma_0, \gamma_1) = (\gamma_{0k} - \bar{\gamma}_0)(\gamma_{1k} - \bar{\gamma}_1),$$

across all values of the index k in the given sub-sample and then estimate the variance of $\tau_k(\xi, \gamma_0)$, $\tau_k(\gamma_0, \gamma_1)$ and $\tau_k(\gamma_0, \gamma_1)$ based on their estimated spectral densities at frequency zero.

Each such comparison is conducted over three individual, separated bifurcations of the full K-length sample: (i) the first 10% of the chain ($L = K/10$) vs. the final 50% ($L = K/2$), (ii) the first 20% of the chain ($L = K/5$) vs. the final 50%, and (iii) the first 30% of the chain ($L = 0.3K$) vs. the final 50%. An individual diagnostic comparison is considered as a 'pass' if the corresponding Z-statistic is less than 1.96 in absolute value. For the two-parameter models, to 'pass' the full diagnostic at each bifurcation, all three measures—mean of ξ, mean of γ_0, and covariance of $\{\xi,\gamma_0\}$—must individually pass; for the three-parameter models, in addition, the mean of γ_1, and the covariances of $\{\xi, \gamma_1\}$ and $\{\gamma_0, \gamma_1\}$ must also pass. The comparisons are

performed in sequential order: if the 10%-vs.-50% diagnostic fails, then the 20%-vs.-50% diagnostic is conducted. If this fails, diagnostic is move to the 30%-vs.-50%. When a diagnostic passes, the indicated early portion of the chain is viewed as a reasonable burn-in. $K_0 < K$ is denoted as the resulting index that begins the retained portion of the chain. So, e.g., if the 10%-vs.-50% diagnostic fails but then the 20%-vs.-50% diagnostic passes, the first 20% of the chain is taken as burn-in and the remaining 80% of the chain is used as the Monte Carlo sample from $\pi(\xi, \gamma_0|\boldsymbol{Y})$ or $\pi(\xi, \gamma_0, \gamma_1|\boldsymbol{Y})$. If K =100,000, as used throughout, this gives K_0 =20,001.

If none of the three sequential diagnostics pass, the random-number generator's seed is re-set and a new set of K Monte Carlo draws is reassessed. If after five such re-starts the diagnostic continues to fail, an *algorithm failure* is reported, and the analysis with the given data set is terminated.

3.7 Bayesian Estimation and Inference on the BMD

The Monte Carlo chain of draws $\{\xi_k\}_{k=K_0}^{K}$ represents an approximation to the marginal posterior for ξ (Gelman *et al.*, 2004, §3). The larger the value of K and the more stable the resulting chain (as determined by the diagnostics mentioned above) are, the better the approximation is. Experience suggests that a value of $K = 100,000$ generally provides stable results. This requires that non-trivial computing resources be brought to bear on the problem, but this is not inconsistent with ongoing practice in contemporary Monte Carlo-based Bayesian calculation (Belloni and Chernozhukov, 2009).

For estimating the BMD, standard decision-theoretic principles are followed and the Bayes estimator is selected based on minimization of the Bayes risk. This requires specification of an underlying loss function. As is well-known, choice of squared-error loss leads to the posterior mean, $\mathrm{E}[\xi|\boldsymbol{Y}]$, as the Bayes estimator, while absolute-error loss leads to the posterior median (Casella and Berger, 2002, §7.3.4). These quan-

tities can be approximated using our our Monte Carlo sample's arithmetic mean, $\frac{1}{K^*}\sum_{k=K_0}^{K}\xi_k$, or after ordering the draws into $\{\xi_{(k)}\}_{k=K_0}^{K^*}$, using the sample's median, $\xi_{([K^*+1]/2)}$ if K^* is odd or $\frac{1}{2}\{\xi_{(K^*/2)}+\xi_{([K^*/2]+1)}\}$ if K^* is even, respectively. If subject-matter considerations cannot guide the choice of loss function, and thus which estimator to employ, appeal to absolute-error loss and the posterior median might be preferable here. This suggestion is, admittedly, empirical: with small sample sizes the (approximated) posterior for ξ is generally found to exhibit a right skew, and the median is more robust against large skew toward higher benchmark dose levels. This has an important, practical consequence: note that a larger estimated BMD implies a higher level of 'acceptable' exposure to a potential toxic agent. If this estimated value is artificially inflated due to anomalies such as a heavy right skew, any public health or environmental guidelines based on the estimate may be unnecessarily lax, and even unsafe. Using the posterior median rather than the posterior mean in such a situation would represent a more-precautionary course of action.

Both the squared-error and absolute-error losses discussed above treat deviations symmetrically, however, an *asymmetric* loss function might better be employed. In effect, a high BMD quantitatively views large exposures to a potentially hazardous agent as relatively safe. If, in truth, the agent is highly toxic, the consequences of such a decision could be severe from a public health or environmental safety perspective. This is more consequential than incorrectly driving BMD $\to 0$ and imposing strict exposure limitations on an innocuous or weakly toxic agent. Unless economic considerations can counterbalance these safety concerns, BMD overestimation generates a greater 'loss'. Therefore the loss function is built asymmetrically and overestimation of the BMD is treated more harshly than underestimation.

Many asymmetric loss constructions are possible; here, selected is a simple bilinear loss function O'Hagan (1994, §2.46):

$$L(\Delta,\xi)=\begin{cases} a(\xi-\Delta) & \Delta\leq\xi \\ b(\Delta-\xi) & \Delta>\xi \end{cases} \tag{3.7.1}$$

where ξ is the target quantity, estimated by the decision Δ. The constants a and b tune the bilinear loss for each individual application. (When $a = b$, the absolute-error loss is recovered.) In our case, overestimation of ξ incurs greater loss than underestimation, so $b > a > 0$ is set. O'Hagan (1994) shows that under (3.7.1), the optimal Bayes estimator for ξ is the $100\left(\frac{a}{a+b}\right)$ th percentile of the posterior distribution $\pi(\xi|\boldsymbol{Y})$. Specific choices for these tuning values will depend on the particulars of the risk-analytic question(s) being addressed; as a default for benchmark dose estimation, it is recommended here to set the ratio $\frac{a}{b} = \frac{1}{2}$, i.e., overestimation of ξ incurs twice as much relative loss as underestimation. If so, the optimal Bayesian estimator for ξ becomes the lower/first tercile of $\pi(\xi|\boldsymbol{Y})$. This is estimated with the lower tercile from the Monte Carlo chain $\{\xi_k\}_{k=K_0}^{K}$, i.e., $\widehat{\xi}_{100\text{BMR}} = \xi_{(\lfloor K^*/3 \rfloor)}$, where $\lfloor x \rfloor$ is the standard floor function which returns the largest integer smaller than x.

Of course, in benchmark analysis the more-critical statistical quantity is the BMDL, since this is the 'point of departure' measure employed in further risk assessments (Kodell, 2005; Izadi *et al.*, 2012). Here, a Bayesian BMDL, say, $\underline{\xi}_{100\text{BMR}}$, is essentially a one-sided, lower, $100(1-\alpha)\%$ credible limit on ξ, satisfying $P(\xi > \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}) = 1 - \alpha$. At the traditional level of $\alpha = 0.05$, what is desired is the lower 5^{th} percentile of $\pi(\xi|\boldsymbol{Y})$, and this is approximated using the lower 5^{th} percentile from our Monte Carlo sample of ξ. Thus the Bayesian BMDL takes the form $\underline{\xi}_{100\text{BMR}} = \xi_{(\lfloor 0.05K^* \rfloor)}$.

3.8 Prior Sensitivity: ϵ-Contamination Analysis

As mentioned previously, experiences with ξ show that this parameter generally varies with a right-skew, which helps motivate the IG prior specification. As mentioned above, however, this choice is open to other options. One obvious alternative is the traditional Gamma distribution: $\xi \sim \mathit{Gamma}(\alpha, \beta)$. Here α and β are elicited using the same expert opinion as detailed previously.

In order to investigate the influence of an alternative prior assumption on the eventual value of $\underline{\xi}_{100\text{BMR}}$, following O'Hagan (1994, §7.15) the prior density on ξ is rewritten as a contaminated mixture of IG and the Gamma prior. Treating the IG density as the base prior, $\pi_0(\xi)$, and the Gamma prior as the contaminating density, $q(\xi)$, the prior for ξ is written as

$$\pi(\xi) = (1-\epsilon)\pi_0(\xi) + \epsilon q(\xi), \tag{3.8.1}$$

where $\epsilon \in [0,1]$ controls the degree of prior contamination. When $\epsilon = 0$, no contamination appears and the prior for ξ is the base IG density; when $\epsilon = 1$, the prior density function for ξ is completely replaced by the contaminating Gamma prior.

Employing the prior density in the ϵ-contamination form from (3.8.1), the estimated BMDL for a specific ϵ is denoted as $\underline{\xi}_{100\text{BMR}}(\epsilon)$. Then, the evolution of $\underline{\xi}_{100\text{BMR}}(\epsilon)$ can be monitored as ϵ increases from 0 to 1.

An additional way to quantify the influence of the ϵ-contaminating prior is to calculate the absolute value of the instantaneous rate of change of the posterior inference at $\epsilon = 0$ (O'Hagan, 1994, §7.15). This is taken as the absolute value of the first derivative of $\underline{\xi}_{100\text{BMR}}(\epsilon)$ at $\epsilon = 0$:

$$|D(q)| = \left| \lim_{\epsilon \to 0} \frac{\underline{\xi}_{100\text{BMR}}(\epsilon) - \underline{\xi}_{100\text{BMR}}(0)}{\epsilon} \right|.$$

O'Hagan showed that this computation simplifies to

$$D(q) = \left[\underline{\xi}_{100\text{BMR}}(1) - \underline{\xi}_{100\text{BMR}}(0)\right] \frac{m_q(\boldsymbol{Y})}{m_0(\boldsymbol{Y})}, \tag{3.8.2}$$

where $m_q(\boldsymbol{Y})$ and $m_0(\boldsymbol{Y})$ are the marginal likelihoods under the contaminating prior density, $q(\xi)$, and base IG prior density, $\pi_0(\xi)$, respectively.

In order to calculate $|D(q)|$, values of $\underline{\xi}_{100\text{BMR}}(1)$ and $\underline{\xi}_{100\text{BMR}}(0)$ are obtained as the estimated BMDLs using $q(\xi)$ and $\pi_0(\xi)$, respectively.

As mentioned in §3.4, the marginal likelihoods, $m_q(\boldsymbol{Y})$ and $m_0(\boldsymbol{Y})$ are intractable integrals. When the AM sample is obtained, however, they can be approximated in

relatively easier ways. [See Ando (2010, §6) for examples.] Here, what is employed is a so-called *geometric estimator* via the bridge sampling method recommended by Meng and Wong (1996) and Lopes and West (2004). Take the two-parameter models for example, it works as follows: An approximation $g(\cdot)$ to the joint posterior density of (ξ, γ_0) is first chosen. For simplicity, $g(\xi, \gamma_0)$ is taken as the bivariate normal density function with mean set to the empirical mean vector and variance set to the empirical variance-covariance matrix of ξ and γ_0, estimated from the AM sample. Next, a set of $K^* = K - K_0$ bivariate vectors $\left\{[\xi_j^* \ \gamma_{0j}^*]^T\right\}_{j=1}^{K^*}$ are drawn from $g(\xi, \gamma_0)$. The marginal p.d.f. is then approximated by substituting the (retained) AM sample, $\{\xi_k, \gamma_{0k}\}_{k=K_0}^{K}$, and the generated sample from $g(\xi, \gamma_0)$, $\left\{[\xi_j^* \ \gamma_{0j}^*]^T\right\}_{j=1}^{K^*}$, into the ratio

$$\widehat{m}(\boldsymbol{Y}) = \frac{\frac{1}{K^*}\sum_{j=1}^{K^*}\left\{\pi\left(\xi_j^*|\alpha,\beta\right)\pi\left(\gamma_{0j}^*|\psi,\omega\right)f\left(\boldsymbol{Y}|\xi_j^*,\gamma_{0j}^*\right)/g\left(\xi_j^*,\gamma_{0j}^*\right)\right\}^{1/2}}{\frac{1}{K^*}\sum_{k=K_0}^{K}\left\{g\left(\xi_k,\gamma_{0k}\right)/\left[\pi\left(\xi_k|\alpha,\beta\right)\pi\left(\gamma_{0k}|\psi,\omega\right)f\left(\boldsymbol{Y}|\xi_k,\gamma_{0k}\right)\right]\right\}^{1/2}},$$

where $f(\cdot)$ denotes the binomial likelihood function and $\pi(\cdot)$ denotes the pertinent prior density [e.g., $q(\xi)$ when approximating $m_q(\boldsymbol{Y})$ or $\pi_0(\xi)$ when approximating $m_0(\boldsymbol{Y})$]. The above gives an example for calculating a marginal likelihood for the two-parameter models. A marginal likelihood for the three-parameter models is calculated in the same fashion by including the $\pi(\gamma_1|\kappa,\lambda)$ prior and using a trivariate approximating normal density $g(\xi, \gamma_0, \gamma_1)$. The geometric bridge sampling method is also used for marginal likelihoods estimations in Bayes factor calculation and posterior model probability calculation (a necessary step for Bayesian model averaging) in later chapters.

Through these steps, the $|D(q)|$ is calculated. It can be used as a measure of sensitivity in the posterior inference as affected by the contaminating prior: the smaller the measure, the less sensitive the posterior inference is in the direction of the contaminating prior (O'Hagan, 1994, §7.15).

Chapter 4
Bayesian Model Averaging

4.1 Model Uncertainty and Model Adequacy

The selection of parametric forms for $R(d)$ presented in Chapter 2 illustrates the wide variety of dose-response functions available to the risk analyst. Many of these operate well at (higher) doses near the range of the observed quantal outcomes; however, they can produce wildly different estimates on BMDs at very small levels of risk (Faustman and Bartell, 1997; Kang *et al.*, 2000). The method described in previous chapters assumed, however, that the choice of dose-response function was made without any uncertainty, i.e., that the specification for $R(d)$—reparameterized or otherwise— was unambiguous and correct. In practice, the extensive library of quantal dose-response forms for $R(d)$ can lead to uncertainty in the model specification. To mitigate concerns over model adequacy/uncertainty, some users have turned to formal model selection procedures for benchmark analysis (Davis *et al.*, 2012). This is an attractive option, although it relies on a reliable selection criterion. In fact, model selection based on the popular Akaike information criteria (AIC) (Akaike, 1973) has been shown to select incorrect models for BMD estimation almost as often as it selects correct models (West *et al.*, 2012). As an alternative to simple model selection methods, and to provide a more model-robust option for BMD estimation, recent work has appeared that employs model averaging techniques. Bayesian model averaging (BMA; see Hoeting *et al.*, 1999) has become popular in this regard, and a number of articles have applied some form of BMA to benchmark analysis; see, e.g., Bailer *et al.* (2005); Morales *et al.* (2006); Shao and Small (2011); Shao (2012); Shao and Gift (2014); Simmons *et al.* (2013); and the references therein.

Suppose $Q > 1$ quantal-response models are under consideration. These form an

uncertainty class U_Q, with individual model elements M_q $(q = 1, \ldots, Q)$. Following Hoeting *et al.* (1999), the model-averaged posterior density for ξ is defined as a mixture of marginal posterior densities for ξ under each model, $f(\xi|\boldsymbol{Y}, M_q)$, with the individual-model posterior probabilities $P(M_q|\boldsymbol{Y})$ employed as weights:

$$f(\xi|\boldsymbol{Y}, U_Q) = \sum_{q=1}^{Q} f(\xi|\boldsymbol{Y}, M_q) P(M_q|\boldsymbol{Y}). \tag{4.1.1}$$

The posterior model probabilities are obtained through Bayes' formula:

$$P(M_q|\boldsymbol{Y}) = \frac{f(\boldsymbol{Y}|M_q)P(M_q)}{\sum_{q=1}^{Q} f(\boldsymbol{Y}|M_q)P(M_q)}, \tag{4.1.2}$$

where $P(M_q)$ are the prior model probabilities. If the risk assessor has no preference for any particular model, a default can be to view the M_qs as equally valid, so that $P(M_q) = 1/Q \quad \forall q$. The posterior model probabilities are then simply $P(M_q|\boldsymbol{Y}) = \frac{f(\boldsymbol{Y}|M_q)}{\sum_{q=1}^{Q} f(\boldsymbol{Y}|M_q)}$. Clearly, the challenge in this sort of Bayesian model averaging (BMA) is to estimate the $P(M_q|\boldsymbol{Y})$ values. This requires accurate estimation of the marginal likelihood $f(\boldsymbol{Y}|M_q)$ which associated integral is seen to be intractable. To approximate it, the geometric bridge sampler of Meng and Wong (1996) is again applied. Details of this algorithm can be found in §3.8.

Given the various components in this multimodel, hierarchial construction, one simple

4.2 Bayesian Model Averaged BMD and BMDL

As pointed out in Hoeting *et al.* (1999), if the square-error loss function is chosen, then the BMA point estimate of ξ may be calculated as a weighted average of the individual-model, point estimates of $\xi_{100\text{BMR}}$ across the uncertainty class:

$$\bar{\xi}_{100\text{BMR}} = \sum_{q=1}^{Q} w_q \hat{\xi}_{100\text{BMR};q}\,, \tag{4.2.1}$$

where the weights are the posterior model probabilities, $w_q = P(M_q|\boldsymbol{Y})$ and the model-specific point estimates of $\xi_{100\text{BMR}}$ are obtained from the posterior mean of each model-specific AM sample. Admittedly, any estimates based on posterior quantiles (e.g., the recommended posterior lower tercile estimates after specifying (3.7.1) with $\frac{a}{b} = \frac{1}{2}$) should not be employed in (4.2.1). Simply put, the quantile of a (weighted) average is not necessary the average of the quantiles. (This is not a concern if the posterior mean is used as the point estimate.) The primary focus in benchmark dose analysis is, however, the BMDL; the BMD point estimator is less important. Thus the BMA point estimator for ξ is not pursued here.

The corresponding Bayesian model-averaged 95% BMDL needs to satisfy $P(\xi \leq \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, U_Q) = 0.05$, where ξ has p.d.f. defined by (4.1.1). This is approximated via

$$P(\xi \leq \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, U_Q) \approx \sum_{q=1}^{Q} w_q \frac{1}{K_q^*} \sum_{j=1}^{K_q^*} I\{\xi_j \leq \underline{\xi}_{100\text{BMR}}|M_q\},$$

where K_q^* is the size of the AM sample from model M_q (after burn-in) and I denotes the indicator function. Here gives a brief proof:

Proof. Equation (4.1.1) defines the model-averaged posterior density for the BMD, $f(\xi|\boldsymbol{Y}, U_Q)$, as a mixture of marginal posterior densities for ξ under each qth model, $f(\xi|\boldsymbol{Y}, M_q)$ using each model's posterior probability $P(M_q|\boldsymbol{Y})$ as the corresponding weight, w_q.

Denote $F(\xi|\boldsymbol{Y}, U_Q)$ as the model-averaged posterior distribution function for the BMD and $F(\xi|\boldsymbol{Y}, M_q)$ as the marginal posterior distribution function for ξ under each model. By integrating both sides of Equation (4.1.1)

$$F(\xi|\boldsymbol{Y}, U_Q) = \sum_{q=1}^{Q} w_q F(\xi|\boldsymbol{Y}, M_q).$$

Now, substitute $\underline{\xi}_{100\text{BMR}}$ into the above equation to find

$$\begin{aligned} P(\xi \leq \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, U_Q) &= F(\underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, U_Q) \\ &= \sum_{q=1}^{Q} w_q F(\underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, M_q) \\ &= \sum_{q=1}^{Q} w_q P(\xi \leq \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, M_q). \end{aligned}$$

Clearly then, the probability, $P(\xi \leq \underline{\xi}_{100\text{BMR}}|\boldsymbol{Y}, M_q)$ can be approximated by the sample proportion $\frac{1}{K_q^*}\sum_{j=1}^{K_q^*} I\{\xi_j \leq \underline{\xi}_{100\text{BMR}}|M_q\}$, where K_q^* is the size of the AM sample under M_q (excluding burn-in) and I denotes the indication function. □

The BMA BMDL is then calculated by finding the root of equation

$$\frac{1}{K_q^*}\sum_{j=1}^{K_q^*} I\{\xi_j \leq \underline{\xi}_{100\text{BMR}}|M_q\} = 0.05.$$

To solve this equation numerically, the `R uniroot` function is used. This employs iterative updating until either an exact root is found or the change in root for one step of the algorithm is less than 2^{-13}. This method directly finds the BMA BMDL through the AM samples under each model, hence, it can be called the *direct method*. Other methods involve multilevel Monte Carlo sampling (first layer determines the number of points drawn from each AM sample based on this sample's corresponding posterior model probability; second layer performs a sampling with replacement from each AM sample). Since sampling with replacement is employed, the estimate is no longer directly obtained from the original AM samples, those methods can be viewed as *indirect methods*. [See Morales *et al.* (2006); Simmons *et al.* (2013) for examples.] Experiences show the numeric solution from the direct method is stable and represents the best guess based on the generated AM samples, therefore, it is employed for the BMA BMDL estimation here.

Chapter 5

Example for hierarchical Bayesian Benchmark analysis

5.1 Cumene Carcinogenicity Data

Cumene, the colloquial name for isopropylbenzene (C_9H_{12}), is a hydrocarbon solvent employed in the production of industrial compounds such as phenol and acetone. Occupational and industrial exposures to the chemical are common, so the U.S. National Toxicology Program (NTP) explored various forms of mammalian toxicity to the chemical (U.S. NTP, 2009). For example, Table 5.1 displays quantal-response data on induction of lung tumors (alveolar/bronchiolar adenomas and carcinomas) in laboratory mice after chronic, two-year, inhalation exposure to cumene.

Table 5.1. Quantal carcinogenicity data: Alveolar/bronchiolar adenomas and carcinomas in female B6C3F$_1$ mice after chronic, two-year, inhalation exposure to cumene (C_9H_{12}). Source: U.S. NTP (2009).

Exposure conc. (ppm), d_i	0	125	250	500
Animals with tumors, Y_i	4	31	42	46
Animals tested, N_i	50	50	50	50

A clear dose response is evidenced. Of additional interest, however, is calculation of a benchmark reference exposure level to inform risk characterization of this potential carcinogen. Notice that the C_9H_{12} exposure dose, d, is actually a concentration (in ppm) here, and so technically we will compute benchmark concentrations (BMCs) based on the quantal carcinogenicity data.

5.2 Benchmark Dose Analysis using Quantal-linear model

To illustrate the hierarchical Bayesian analysis, the reparameterized quantal-linear model (M3) is first employed (see §2.5 for its reparameterized form). This choice is not trivial, actually, the quantal-linear form $R(d) = 1 - e^{-\beta_0 - \beta_1 d}$ from Table 2.1 is a popular construct for dose-response modeling. Also known as the one-stage model or the complementary-log model, it is a special case of the famous 'multi-stage' model in carcinogenicity testing (Piegorsch and Bailer, 2005, §4.2.1). This model suggests that a sequence of multiple distinct genetic events precede the onset of cancer (Armitage and Doll, 1954). Because of this, it has wide application in toxicological risk assessment with quantal data (Gaylor, 1998; Foronda *et al.*, 2007; Buckley *et al.*, 2009; Shao and Small, 2012).

Next, the BMR is set to the default level of 0.10 (U.S. EPA, 2012). Using input from domain experts, information for prior elicitation was based on existing background in the toxicological literature. Began with ξ: an oral No Observed Adverse Effect Level (NOAEL) for long-term cumene exposure in rodents was given by the EPA IRIS website (`http://www.epa.gov/iris/subst/0306.htm`) as 110 mg/kg-day. Guidance for converting the oral mg/kg-day metric to inhalation ppm (as in Table 5.1) was available via conversion data given in California OEHHA (2004): for rodents, 1 mg/kg-day works out to roughly 2.25 ppm. Arguing that a NOAEL is roughly similar to a BMD—although, see (Kodell, 2005) for a more-focused discussion—the median prior estimate was $Q_{2\xi}$ = 110 mg/kg-day = 247.5 ppm $\approx$ 250 ppm. For the first quartile, the EPA IRIS site gave a cumene inhalation NOAEL for rodents as 435 mg/cu.m, based on a shorter, 13-week study. The shorter exposure period made the toxicity estimate somewhat less-certain when comparing to longer-duration exposures, and was better suited as a conservative estimate on the BMD's central tendency; we translated this to serve as a prior estimate of the first quantile. For conversion to inhalation ppm, the EPA indicated that roughly 1 mg/cu.m = 0.204

ppm, so we took $Q_{1\xi} = 435$ mg/cu.m $= 88.75$ ppm ≈ 90 ppm. We then divided the ppm dose by 500 ppm to standardize the dose scale to $0 \leq d \leq 1$. This produced prior first quartile and median estimates for $\pi(\xi|\alpha, \beta)$ as 0.18 and 0.5, respectively.

For γ_0, background risk data on alveolar/bronchiolar tumors in B6C3F$_1$ mice were given by U.S. NTP (2009, Table D3a). The data indicated a median background response of $Q_{2\gamma} = 0.08$ and a lower quartile of $Q_{1\gamma} = 0.04$. Unfortunately, since there was no evidence of carcinogenic effects of cumene had been evidenced in laboratory mice prior to the NTP's toxicology study (U.S. NTP, 2009), no elicitation for γ_1 was available. As a result, the objective prior $\gamma_1 \sim Beta(\frac{1}{2}, \frac{1}{2})$ is set. Note, M3 is a two-parameter model, therefore, specifying a prior for γ_1 is not necessary here, however, it is still a necessary step for other three-parameter models which will be employed later.

Based on these prior-information summaries, the elicited prior distributions are taken as $\xi \sim IG(0.53, 0.13)$ and $\gamma_0 \sim Beta(1.36, 12.31)$, using the methods described in §3.2. And then, the AM approach is applied to the consequent hierarchical model for $\boldsymbol{\theta} = (\xi, \gamma_0)$. No data failure nor algorithm failure was evidenced with the cumene data. The generated chain passed the Geweke convergence diagnostic at the first 10% level, and so the burn-in was chosen to be the initial $10,000$ iterations. Figure 5.1 displays the trace plot for the ξ component of the AM sample. The plot exhibits stable performance and fairly fast mixing. Figure 5.2 displays a histogram (with overlayed kernel density estimate) for the ξ component of the AM sample. Both suggest a unimodal and slightly right-skewed marginal posterior distribution for ξ.

Figure 5.3 displays the estimated risk function $\hat{R}(d)$ based on the median posterior estimates for (ξ, γ_0), the lower tercile estimates for (ξ, γ_0) and the MLEs $\hat{\xi}$, $\hat{\gamma}_0$, along with the original data proportions. All these curves give reasonably good estimates for $R(d)$, relative to the data, although the median-based Bayesian estimate shifts consistently to the right of the other two risk functions at most levels of d.

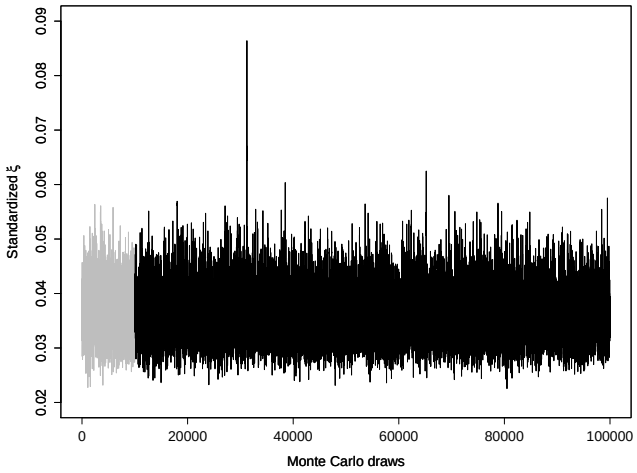

FIGURE 5.1. Trace plot for ξ in Example 1. Dose scale (ξ) is standardized to unit length after dividing by the highest dose in the data set. First 10,000 Monte Carlo draws in grey indicate burn-in period.

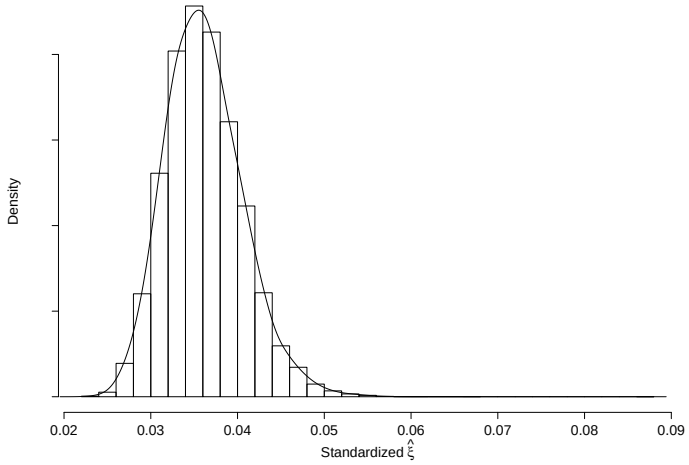

FIGURE 5.2. Histogram and Gaussian kernel density estimator for ξ in Example 1. Doses are standardized to unit length after dividing by the highest dose (500 ppm) in the data set.

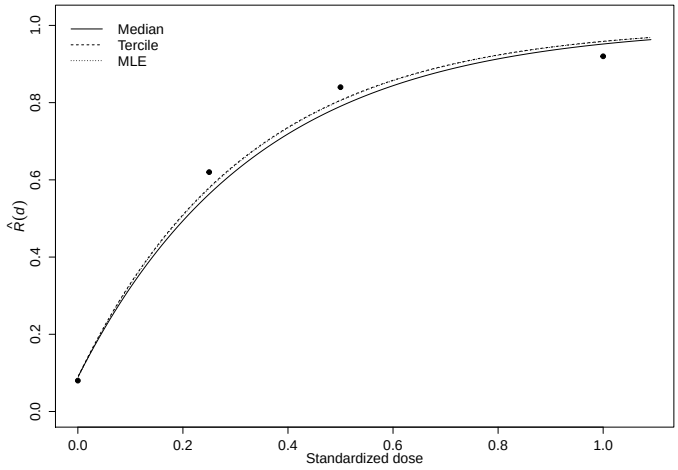

FIGURE 5.3. Estimated risk functions for cumene carcinogenicity data in Example 1. Solid curve (———) is based on the median posterior estimate, dashed curve (- - - - -) is based on the lower tercile posterior estimate, and dotted curve (·····) is based on the maximum likelihood estimate (MLE). The estimated risk function based on the lower tercile posterior estimate is indistinguishable from the estimated risk function based on the MLE at this scale. Solid circles are observed proportion data. Doses are standardized to unit length after dividing by the highest dose (500 ppm) in the data set.

For reporting purposes, the final BMD estimates on the original ppm scale are $\hat{\xi}_{10}$=17.973 ppm (if using the sample posterior median) or $\hat{\xi}_{10}$=17.046 ppm (if using the sample posterior lower tercile). The 95% BMDL is $\underline{\xi}_{10}$=14.752 ppm. Comparing these to frequentist estimates, the MLE is $\hat{\xi}_{10}$=17.062 ppm and a 95% frequentist Wald BMDL is $\underline{\xi}_{10}$=13.618 ppm. Both sets of values rest in similar ranges, and provide comparable points of departure for conducting further risk-analytic calculations on cumene carcinogenicity.

Thus for these data, the Bayesian approach operates similarly to the frequentist analysis, but it also provides an additional benefit for risk assessors: the Bayesian strategy combines elicited prior information on both the background risk and the BMD, potentially improving the estimation process by incorporating more-complete prior information. Besides, the frequentist BMDL, such as the Wald lower limit,

often depends on asymptotic arguments to justify application to small samples. Since the Bayesian BMDL is estimated directly from the distribution of ξ, no asymptotic assumptions are involved.

5.3 Monte Carlo standard error for BMDL

Given the same input data $\boldsymbol{Y}$, the BMDL estimates from the AM chain will vary if different random seeds are assigned. This variation may be quantified via the *Monte Carlo Standard Error*, MCSE. If an AM chain has converged to the target posterior distribution, then a longer chain represents a bigger sample. One then expects that the larger sample will produce a smaller MCSE for the BMDL. MCSE is often estimated via the batch mean method or via spectral density estimation (Flegal and Jones, 2010). By estimating the MCSE for the quantity of interest, one can estimate the length of the simulating chain that leads to an acceptable level of accuracy (Barnes, 2012).

Here, a simple approach is used to estimate the MCSE of the BMDL $\underline{\xi}_{100\text{BMR}}$. After choosing a specific model, a variety of chain lengths ranging from 5,000 to 200,000 are considered. One hundred realizations of $\underline{\xi}_{100\text{BMR}}$ calculated at each chain length and the sample standard deviation of these 100 estimates is calculate to estimate the MCSE of $\underline{\xi}_{100\text{BMR}}$. Figure 5.4 shows the results when fitting the elicited, reparameterized, quantal linear model (M3) to the cumene data. The circles represents the estimated MCSEs of $\underline{\xi}_{100\text{BMR}}$ at each chain length. An overlaid curve smooths these via second-order LOESS estimation. Not unexpectedly, the MCSE decreases as chain length increases. One sees that a rapid decrease in MCSE is followed by a more shallow decrease onece the chain length moves past 75,000. This suggests that 75,000 is long enough to produce an acceptable MCSE of $\underline{\xi}_{100\text{BMR}}$ for these data with this model. The LOESS estimate of MCSE at a length of 75,000 is about 0.00007 which is roughly 0.2% of the estimated $\underline{\xi}_{100\text{BMR}}$ (a standardized BMDL was estimated as

0.0295 by using model M3 with chain length of 100,000). Notice also that by choosing a more conservative chain length of 100,000 (as was seen in the cumene example), the estimated MCSE of $\underline{\xi}_{100\text{BMR}}$ should be slightly below this value of 0.00007.

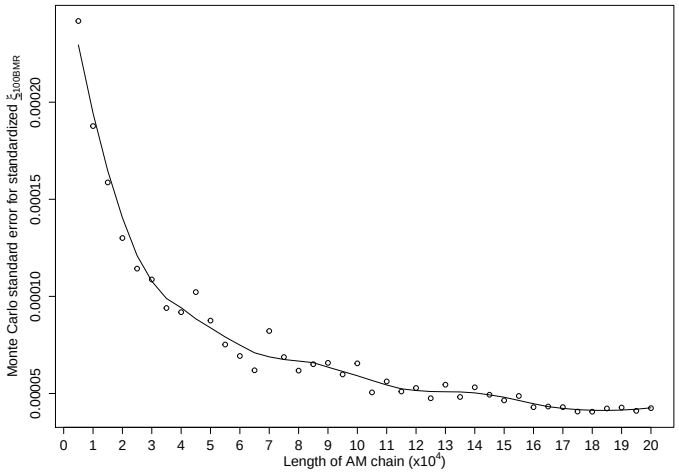

FIGURE 5.4. Monte Carlo standard errors of the standardized $\underline{\xi}_{100\text{BMR}}$ as a function of changing chain length with cumene carcinogenicity data from Table 5.1 and M3 fit. Circles (∘) represents the estimated Monte Carlo standard error of $\underline{\xi}_{100\text{BMR}}$ at each chain length. Solid curve (———) represents the second-order LOESS estimation.

5.4 Prior sensitivity

To explore sensitivity of the IG prior within the benchmark dose analysis for these data, the following analysis returns to the methods in §3.8 and contaminates the base prior for ξ via three scenarios. In scenario 1, an objective $IG(0.001, 0.001)$ form is chosen as the base prior for ξ and is contaminated by the similar, objective, $Gamma(0.001, 0.001)$ prior with p.d.f. $q(\xi) = \frac{0.001^{0.001}}{\Gamma(0.001)}\xi^{0.001-1}e^{-0.001\xi}$. This scenario is used to investigate the sensitivity of the $IG(0.001,\ 0.001)$ prior for $\underline{\xi}_{100\text{BMR}}$ when no elicitation is available. In scenario 2, the elicited IG prior is chosen as the base prior for ξ and is contaminated by a similarly elicited Gamma prior. The same quartile

information for ξ is used to build both elicited priors. Scenario 2 is used to investigate the sensitivity of the IG prior when elicitation is available. In scenario 3, the elicited IG prior is chosen as the base prior for ξ and is contaminated by an objective $Gamma(0.001, 0.001)$ prior. Scenario 3 is used to investigate the robustness of $\underline{\xi}_{100\text{BMR}}$ when the elicited IG prior is contaminated by an objective Gamma prior. Throughout, the prior for γ_0 is taken as either the elicited Beta prior, $Beta(1.36, 12.31)$, or an objective $Beta(\frac{1}{2}, \frac{1}{2})$ prior. This gives six different settings for study, under which the consequent BMDL $\underline{\xi}_{100\text{BMR}}$ is monitored as ϵ increased from 0 to 1. The corresponding $|D(q)|$ instantaneous change measures from (3.8.2) are also calculated and compared. Figure 5.5 displays the consequent evolution of $\underline{\xi}_{100\text{BMR}}$ across these six settings.

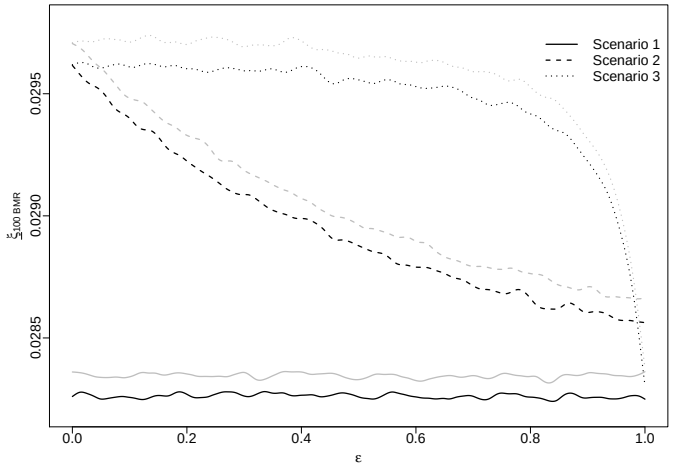

FIGURE 5.5. Kernel smoothed value of $\underline{\xi}_{100\text{BMR}}$ as a function of changing ϵ in ϵ-contamination study with cumene carcinogenicity data from Table 5.1. Solid curves (———) from Scenario 1: an objective $IG(0.001, 0.001)$ prior for ξ is contaminated by an objective $Gamma(0.001, 0.001)$ prior. Dashed curves (- - - - -) from Scenario 2: the elicited IG prior for ξ is contaminated by the elicited Gamma prior. Dashed-dotted curves (·····) from Scenario 3: the elicited IG prior for ξ is contaminated by an objective $Gamma(0.001, 0.001)$ prior. Gray curves indicate an objective $Beta(\frac{1}{2}, \frac{1}{2})$ prior for γ_0; black curves indicate the elicited Beta prior for γ_0. Doses are standardized to unit length after dividing by the highest dose (500 ppm) in the data set.

The three choices for ξ are distinguished using different line types (see the figure legend), and the two Beta priors are distinguished using grey or black coloration. In scenario 1, no substantive change in $\underline{\xi}_{100\text{BMR}}$ is evidenced as ϵ varies from 0 to 1 in Figure 5.5. This suggests that the objective $IG(0.001, 0.001)$ prior is essentially equivalent to an alternative objective $Gamma(0.001, 0.001)$ prior in approximating the improper reciprocal prior, at least for these data.

In scenario 2, $\underline{\xi}_{100\text{BMR}}$ decreases in a roughly linear fashion, to a maximum drop of about 4% as ϵ increased from 0 to 1 (Figure 5.5). This linear decrease suggests that the contaminating Gamma prior affects $\underline{\xi}_{100\text{BMR}}$ in a consistent fashion as ϵ changes. However, the relatively small maximum decrease also suggests that the sensitivity of $\underline{\xi}_{100\text{BMR}}$ to the contaminating elicited Gamma is minor.

In scenario 3, slight decreases in $\underline{\xi}_{100\text{BMR}}$ are evidenced as ϵ increased from 0 to approximately 0.8; then $\underline{\xi}_{100\text{BMR}}$ drops precipitously as ϵ increases closer to 1 (Figure 5.5). However, the maximum drop is roughly 4%, essentially the same as in Scenario 2. This provided some evidence for the robustness of $\underline{\xi}_{100\text{BMR}}$ under an elicited IG prior; a significant amount of contamination from the objective prior is required to greatly reduce the BMDL. In each scenario, also seen is that these patterns occur for either form of prior for γ_0, although, on average, the elicited prior consistently produces a roughly 0.3% smaller BMDL than the objective Beta prior.

Table 5.2 gives the values for $|D(q)|$ under each scenario. Also included is a relative error, measuring the maximum change of $\underline{\xi}_{100\text{BMR}}$ compared to $\underline{\xi}_{100\text{BMR}}(0)$:

$$\delta_{100\text{BMR}} = \frac{\underline{\xi}_{100\text{BMR}}(0) - \min\left(\underline{\xi}_{100\text{BMR}}(\epsilon)\right)}{\underline{\xi}_{100\text{BMR}}(0)}.$$

The relative errors for scenario 1 are all less than 1%, suggesting little overall decrease in $\underline{\xi}_{100\text{BMR}}$ as ϵ increases from 0 to 1. For scenarios 2 and 3, the relative errors are all approximately 4%. Although non-trivial, such small changes suggest only minor sensitivity of the base priors to the contaminating priors, at least for

TABLE 5.2. Relative errors, $\delta_{100\text{BMR}}$, and instantaneous change measure, $|D(q)|$, for each prior contamination scenario (see text) with the cumene carcinogenicity data in Table 5.1.

	Objective Beta prior		Elicited Beta prior	
	$\delta_{100\text{BMR}}$	$\lvert D(q)\rvert$	$\delta_{100\text{BMR}}$	$\lvert D(q)\rvert$
Scenario 1	7.679×10^{-3}	1.134×10^{-4}	5.287×10^{-3}	5.632×10^{-6}
Scenario 2	3.767×10^{-2}	1.905×10^{-3}	3.645×10^{-2}	1.839×10^{-3}
Scenario 3	4.477×10^{-2}	4.536×10^{-5}	4.396×10^{-2}	4.174×10^{-5}

these data. Values of $|D(q)|$ for scenario 1 and 3 are all very close to 0 suggesting tiny instantaneous change of $\underline{\xi}_{100\text{BMR}}$ at $\epsilon = 0$. These are consistent with the graphical patterns seen in Figure 5.5. By contrast, for scenario 2 $|D(q)|$ is roughly 1 to 3 orders of magnitudes higher than the other scenarios. This is again consistent with the graphical patterns in Figure 5.5.

5.5 Model Uncertainty and BMA

The previous representation for the risk function was based on the popular quantal linear model. As mentioned earlier, many other forms can be chosen to model $R(d)$, however. For instance, the logistic form $R(d) = 1/(1 + e^{-\beta_0-\beta_1 d})$ from Table 2.1 is another highly popular construct for dose-response modeling (Foronda *et al.*, 2007; Shao and Small, 2012). Employing the same elicitation strategies for the priors and the AM approach for posterior approximation. If the reparameterized logistic model is applied to the cumene carcinogenicity data in Table 5.1, the BMD estimates on the original ppm scale are $\hat{\xi}_{10}$=42.946 ppm (if using the sample posterior median) or $\hat{\xi}_{10}$=40.892 ppm (if using the sample posterior lower tercile). No 'algorithm failure' has been encountered and the burn-in is found to be $10,000$ based on the convergence diagnostic test. The 95% BMDL is $\underline{\xi}_{10}$=35.599 ppm. These values contrast greatly

with those from the quantal linear model in §5.2, leading one to ask which is more appropriate—a classic example of model uncertainty. To compare the two fits in Bayesian framework, the Bayes Factor (BF) (Kass and Raftery, 1995) can be used: the BF comparing the quantal linear model to the logistic model is calculated as the ratio of corresponding marginal likelihoods:

$$BF = \frac{m_{\text{ql}}(\boldsymbol{Y})}{m_{\text{lo}}(\boldsymbol{Y})}\,, \tag{5.5.1}$$

where $m_{\text{ql}}(\boldsymbol{Y})$ denotes the marginal likelihood using the quantal linear model and $m_{\text{lo}}(\boldsymbol{Y})$ denotes the marginal likelihood using the logistic model. Both marginal likelihoods are approximated using the geometric estimator and bridge sampling method from §3.8. Doing so for the cumene data produces $BF = 518.3$. Following Kass and Raftery (1995, §3.2), since this BF exceeds 150, it suggests that there is 'very strong' evidence that the quantal linear model fits the data better than the logistic model. (This is, perhaps, not surprising: the concave response observed in Figure 5.3 would be difficult for a logistic model to match, compared to the always-concave quantal linear form.)

The above illustrated the model uncertainty issue in the benchmark dose analysis regarding to cumene data and a simple model comparison using Bayes Factor suggested using the quantal linear model instead of the logistic model. To illustrate the hierarchical BMA approach with the eight reparameterized models in §2.2, with the same prior structures in place (Note, here for the three-parameter models, $\gamma_1 \sim Beta(\frac{1}{2}, \frac{1}{2})$ is specified additionally.), a Monte Carlo AM sample for the remaining six models (samples for quantal linear and logistic have been obtained) is generated. Like what has been seen in §5.2 and above, no 'algorithm failures' have been encountered, and the six newly generated chains all passed the convergence diagnostic tests. This produced burn-ins of $10,000$ initial iterations for each chain, allowing to operate with $90,000$ AM draws for each model. Figures 5.6 and 5.7 dis-

play the estimated risk functions $\hat{R}(d)$ along with observed risks for all 8 models. In these reparameterized models, ξs are based on posterior lower tercile, γ_0s are based on posterior median and γ_1s for three-parameters models are also based on posterior median. These plots reflect how well each model fits the data. A close comparison between them leads to conclude: M3, M6 and M7 give the best fit, M5 and M8 are after them, while M1, M2 and M4 fit poorly.

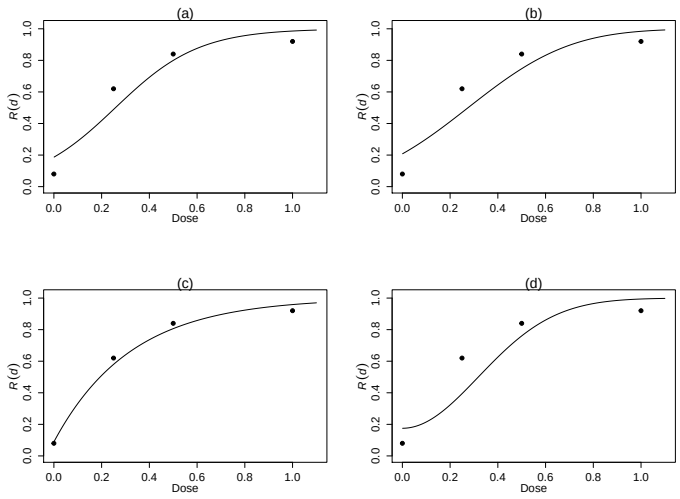

FIGURE 5.6. Estimated risk functions from models M1 to M4 for cumene carcinogenicity data in Example 1. (a) is M1, (b) is M2, (c) is M3 and (d) is M4. Estimating curves are based on the lower tercile posterior estimate. Solid circles are observed proportion data. Doses are standardized to unit length after dividing by the highest dose (500 ppm) in the data set.

The model-specific, lower-tercile benchmark concentrations (BMCs) are reported in Table 5.3. The table also lists the corresponding 95% model-specific BMCLs, along with the posterior model probabilities/weights, $w_q = P(M_q|\boldsymbol{Y})$, assuming uniform prior model probabilities for each model. The 95% BMA BMCL is calculated by using the 'direct method' as illustrated in §4.2, which produces $\underline{\xi}_{10} = 15.1927$ ppm. These are also reported at the bottom of Table 5.3.

From the table, it's seen that the posterior model probabilities vary widely with

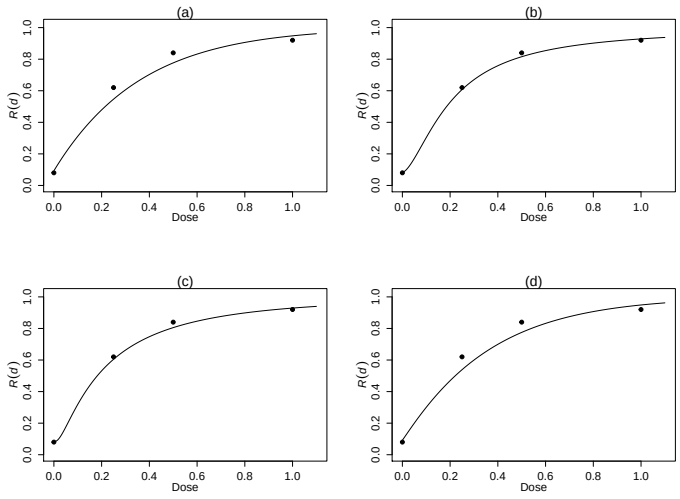

FIGURE 5.7. Estimated risk functions from models M5 to M8 for cumene carcinogenicity data in Example 1. (a) is M5, (b) is M6, (c) is M7 and (d) is M8. Estimating curves are based on the lower tercile posterior estimate. Solid circles are observed proportion data. Doses are standardized to unit length after dividing by the highest dose (500 ppm) in the data set.

these data, however, suggesting that certain models may provide better-quality estimates than others. The three-parameter log-logistic and log-probit give the highest weights, both near 35%, followed by the two-parameter quantal linear model at 23%. All other models show posterior probabilities below 5%.

The BMA 95% BMCL is 15.1927 ppm, within the range of the individual model BMCLs, and closer to the individual lower limits associated with higher-weight models. The ramifications for the risk analyst here are substantial: had a choice of the poorly-fitting logistic, probit, or quantal-quadratic models been made for analyzing these data, the consequent BMC and BMCL would have been far too large. By integrating information across the various models, however, a more-tempered, model-robust estimate is produced from which further risk analytic calculations on cumene carcinogenicity can be conducted. Corroborating reports by many who have come before, we find that BMA adjustment frees the risk assessor from the selection biases,

TABLE 5.3. BMC estimates based on posterior lower terciles and 95% BMCLs (in ppm) from each reparameterized model in §2.2, along with corresponding Bayesian model averaged (BMA) BMDL, for cumene carcinogenesis example. The BMR is set to 0.10.

Model	BMC_{10}	$BMCL_{10}$	$P(M_q\|\boldsymbol{Y})$
M_1	40.8915	35.5991	0.00044
M_2	42.5753	37.6845	0.00005
M_3	17.0118	14.7567	0.22887
M_4	72.9964	66.2304	0.00000
M_5	19.1450	16.2568	0.01356
M_6	25.7309	15.9229	0.36905
M_7	24.7242	15.3244	0.34956
M_8	21.0055	17.0606	0.03846
BMA		15.1927	

model inadequacies, and inferential uncertainties one encounters when committing to only a single parametric model to perform the benchmark analysis.

Chapter 6

Performance Evaluations

To explore the features of the reparameterized Bayesian hierarchical model in further detail, a series of Monte Carlo evaluations was performed. The primary goals are four folds: 1. study the stability of the method in terms of algorithm failures for different configurations and sample sizes. 2. study the asymptotic property for the point estimate of model-specific BMD. 3. study the empirical coverage rates for the the model-specific BMDLs. 4. study the characteristics for the Bayesian model averaged BMDL estimate.

6.1 Simulation design

Throughout, the BMR was set to the typical default level of BMR = 0.10 (U.S. EPA, 2012) and operated at a credible level of 95%. The doses were taken at four levels: $d_1 = 0$, $d_2 = 0.25$, $d_3 = 0.5$, $d_4 = 1$, corresponding to a standard design in cancer risk experimentation (Portier, 1994). Equal numbers of subjects, $N_i = N$, were taken per dose group. Three different per-dose sample sizes were considered: $N = 25$, 50, or 1000; the latter approximates a 'large-sample' setting, while the former two are more commonly seen in, e.g., environmental carcinogenicity investigations (as with the cumene data in Example 1). As throughout, all of the calculations were performed within the `R` programming environment (R Development Core Team, 2012).

For the true dose-response model the eight reparameterized dose-response models listed in Section 2.2 were adopted. The background risk, γ_0, was set between 1% and 30%, and the other risk levels were increased to produce a variety of (strictly) increasing forms, ending with high-dose risks at $d = 1$ between 10% and 90%. To set the parameters for the two-parameter models, $R(d)$ were first fixed at $d = 0$ and

$d = 1$ to produce two equations ($\gamma_0 = R(0)$ was given directly), and the unknown ξ was solved. To set the parameters for the three-parameter models, $R(d)$ were fixed at $d = 0$, $d = 1/2$ and $d = 1$ to establish 3 equations (here, $\gamma_0 = R(0)$ and $\gamma_1 = R(1)$ were given directly) and the unknown ξ was solved. The actual specifications for $R(0)$, $R(1)$ and $R(1/2)$ and resulting parameter configurations for the various models are given in Table 6.1.

TABLE 6.1. Models and configurations for the Monte Carlo evaluations. Model codes are from Table 2.1

Model	Parameterization	C1	C2	C3	C4	C5
	$R(0)=$	0.01	0.10	0.05	0.30	0.10
	$R(1)=$	0.10	0.30	0.50	0.75	0.90
M1	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0401	0.5535	0.3974	0.1619	0.1700
M2	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0476	0.5331	0.3567	0.1606	0.1575
M3	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.1054	0.4192	0.1642	0.1023	0.0480
M4	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0514	0.6475	0.4052	0.3199	0.2190
	$R(0)=$	0.01	0.10	0.05	0.30	0.10
	$R(\frac{1}{2})=$	0.04	0.17	0.30	0.52	0.50
	$R(1)=$	0.10	0.30	0.50	0.75	0.90
M5	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0602	0.5911	0.1783	0.1818	0.1925
	γ_1	0.1000	0.3000	0.5000	0.7500	0.9000
M6	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0648	0.5848	0.2083	0.2439	0.2760
	γ_1	0.1000	0.3000	0.5000	0.7500	0.9000
M7	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0711	0.5789	0.2267	0.2608	0.2794
	γ_1	0.1000	0.3000	0.5000	0.7500	0.9000
M8	γ_0	0.0100	0.1000	0.0500	0.3000	0.1000
	ξ	1.0634	0.5874	0.1852	0.2072	0.2025
	γ_1	0.1000	0.3000	0.5000	0.7500	0.9000

Then, for each of the 8 (models) $\times$ 5 (configurations) $\times$ 3 (sample sizes) = 120 combinations, 2000 pseudo-binomial quantal-response data sets were simulated via `R`'s `rbinom` function. And then for each data set, eight AM samples were generated, one from each of the eight models. Different model-specific point estimates for the BMD $\hat{\xi}_{10;q}$ were calculated using the mean, median and lower-tercile. The 95% BMDL $\underline{\xi}_{10;q}$ was taken as the lower 5^{th} percentile of each model's AM sample. The Bayesian model-averaged BMDL was calculated by using the 'direct method' mentioned in the same section. For simplicity, no prior elicitation was applied in the simulations and so in the calculations only objective priors were employed for ξ, γ_0 and γ_1, i.e., $\xi \sim IG(0.001, 0.001)$, $\gamma_0 \sim Beta(\frac{1}{2}, \frac{1}{2})$ and $\gamma_1 \sim Beta(\frac{1}{2}, \frac{1}{2})$ to represent parameter uncertainty.

6.2 Bimodal distribution for model-specific point estimates of ξ

As mentioned above, for each model and configuration, there is only one true BMD. Therefore, it's reasonable to believe that any model-specific point estimator of BMD should distribute around that true BMD. Besides, Chao (1970) showed Bayes estimators are asymptotically equivalent to MLE for all priors and loss functions in a certain class. This means the Bayes estimators should exhibit the same asymptotic property as the MLE, such as the asymptotic normality property. During this research, however, some histograms for the three-parameter-model point estimates of ξ exhibited bimodal pattern, even when $N = 1000$. This artifact happened only under configuration code 'C1', the shallowest (also flattest) dose-response configuration.

As shown in Figure 6.1, the two modes in the histogram seem to be separated by $\xi = 1$. This phenomenon motivates a closer investigation at dose level 1. In the simulated data, $d_m = 1$ is the highest dose level. By checking the reparameterized forms of the three-parameter models in 2, if γ_1 is used by default, parameter ξ will

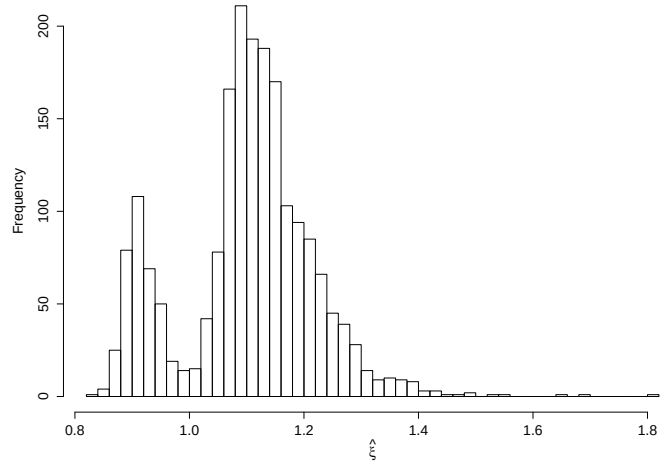

FIGURE 6.1. Histogram for 1977 lower-tercile point estimates of ξ for M7 data with M7 fit using configuration C1, $N = 1000$ and true BMD $= 1.071$. A clear bimodal pattern appears while the two modes seem to be separated by $\xi = 1$ (at the highest dose level).

closely interact with d_m. Indeed, different forms of three-parameter risk functions would be produced when the sign of $\xi - d_m$ changes (see the reparameterized forms in Chapter 2). As one can imagine, this sign will be frequently changed during Monte Carlo updating when the true BMD is close to d_m; this results in two shapes of risk functions and hence two estimates of ξ. Therefore, it seems to suggest that this artifact is related to the situation when the true BMD is very close to d_m. By checking with Table 6.1, this is exactly the case for C1 data.

The goal is hence to avoid close interaction between ξ and a nuisance quantity in the reparameterized model. In cases where the BMD is felt to be near the highest dose—as determined, e.g., from elicited prior information—it's recommended that revising the reparameterization for γ_1 to define it as the response at some lower dose, say, $R(d_2)$. This can ameliorate the instabilities that occur when $\hat{\xi}$ is near d_m. Therefore, for all simulated data generated by three-parameter models with configuration code 'C1', the $R(d_2) = R(0.25)$ is used as a parameter instead of γ_1.

Since $d_2 = 0.25$ which is far less than the true BMD ≈ 1, it's now less likely to interfere the AM algorithm to generate a reasonable sample for ξ.

The structure for the reparameterized three-parameter models is not greatly changed; the only change would be that the close interaction is now between ξ and d_2. These forms are illustrated as follows:

$$\mathbf{M_5}\text{:}\ R(d) = \gamma_0 + (1-\gamma_0)\left[1 - \exp\left\{\frac{C_5 d_2 d(d_2-d) + \Gamma_5 \xi d(\xi - d)}{\xi d_2(\xi - d_2)}\right\}\right],$$
$$\text{where } \Gamma_5 = \log\left(\frac{1-R(0.25)}{1-\gamma_0}\right) \text{ and } C_5 = -\log(1-\text{BMR}).$$

$$\mathbf{M_6}\text{:}\ R(d) = \gamma_0 + (1-\gamma_0)\left[1 + \exp\left\{\frac{C_6[\log d_2 - \log d] + \Gamma_6[\log \xi - \log d]}{\log \xi - \log d_2}\right\}\right],$$
$$\text{where } \Gamma_6 = \log\left(\frac{1-R(0.25)}{1-\gamma_0}\right) \text{ and } C_6 = \log\left(\frac{\text{BMR}}{1-\text{BMR}}\right).$$

$$\mathbf{M_7}\text{:}\ R(d) = \gamma_0 + (1-\gamma_0)\Phi\left(\frac{C_7[\log d_2 - \log d] + \Gamma_7[\log d - \log \xi]}{\log d_2 - \log \xi}\right),$$
$$\text{where } \Gamma_7 = \Phi^{-1}\left(\frac{R(0.25)-\gamma_0}{1-\gamma_0}\right) \text{ and } C_7 = \Phi^{-1}(\text{BMR}).$$

$$\mathbf{M_8}\text{:}\ R(d) = \gamma_0 + (1-\gamma_0)\left(1 - \exp\left\{-\exp\left(\frac{C_8[\log d_2 - \log d] + \Gamma_8[\log d - \log \xi]}{\log d_2 - \log \xi}\right)\right\}\right),$$
$$\text{where } \Gamma_8 = \log\left(-\log\left(\frac{1-R(0.25)}{1-\gamma_0}\right)\right) \text{ and } C_8 = \log(-\log(1-\text{BMR})).$$

6.3 Data Failures and Algorithm Failures

To examine the robustness of the algorithm, the number of 'data failures' and 'algorithm failures' were collected from each configuration. In general, some data failures do occur for the flatter configurations (C1 and C2) at smaller sample sizes. Algorithm failures also happen more for configurations (C1 and C2) while two-parameter models have fewer algorithm failures than three-parameter models. This is not unexpected as three-parameter models are more complicated. This corroborates results seen with similar models in Piegorsch *et al.* (2013).

Table 6.2 illustrates the number of data failures for each of the 8 (models) $\times$ 5 (configurations) $\times$ 3 (sample sizes) = 120 combinations. 'Data failure' is a characteristic for the data and is determined by using a model-free screen method mentioned in §3.5, therefore the number of 'data failures' will not vary for a given data set.

TABLE 6.2. Numbers of data failure for simulated data generated by 8 model, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	116	112	81	116	104	107	92	107
	50	39	34	21	39	33	33	23	33
	1000	0	0	0	0	0	0	0	0
C2	25	58	55	49	72	64	65	56	65
	50	8	7	5	10	8	9	7	9
	1000	0	0	0	0	0	0	0	0
C3	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

Table 6.3 illustrates the numbers of algorithm failures when fitting a logistic model to the simulated data. Tables 6.4 and 6.5 illustrate the numbers of algorithm failures when fitting two-stage model (M5) and log-probit model (M7) to the simulated data, respectively. Algorithm failures regarding to other models are shown in Appendix A. As mentioned in §6.2, all two-parameter models used the default (ξ, γ_0) parameterization, however, three-parameter models used $(\xi, \gamma_0, R(0.25))$ parameterization for configuration code C1 and default $(\xi, \gamma_0, \gamma_1)$ for the other configurations.

Different from the numbers of data failures, the numbers of algorithm failures can vary within a data set. A different parameterization may reduce the numbers of algorithm failures, for example, if the $(\xi, \gamma_0, R(0.25))$ parameterization is used for three-parameter models on configuration C1 data, there will be fewer algorithm failures. Besides, if a different random seed is used, the numbers of algorithm failures will be slightly different too. The displayed numbers, however, reflect the general algorithm failure pattern. In addition, algorithm failure seems to be irrelevant to

TABLE 6.3. Numbers of algorithm failures when fitting logistic model (M1) to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	23	26	25	28	23	35	20	23
	50	6	14	16	7	9	16	10	11
	1000	0	0	0	0	0	0	0	0
C2	25	11	8	18	17	8	12	11	15
	50	8	2	11	5	5	5	6	4
	1000	0	0	0	0	0	0	0	0
C3	25	0	1	0	0	0	1	1	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	3	3	3	0	2	1	0	2
	50	0	0	1	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

TABLE 6.4. Numbers of algorithm failures when fitting two-stage model (M5) to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	22	21	44	14	28	21	41	29
	50	7	9	7	3	6	13	4	3
	1000	0	0	0	0	0	0	0	0
C2	25	109	93	109	86	95	95	95	98
	50	46	43	56	42	41	52	42	50
	1000	0	0	0	0	0	0	0	0
C3	25	1	3	0	1	0	2	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	1	1	1	2	0	1	1	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	1	0

TABLE 6.5. Numbers of algorithm failures when fitting log-probit model (M7) to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	79	118	117	87	109	90	91	101
C1	50	47	50	44	43	45	50	45	56
C1	1000	0	0	0	0	0	0	0	0
C2	25	28	36	34	61	41	34	38	42
C2	50	28	22	24	25	32	30	34	32
C2	1000	0	0	0	0	0	0	0	0
C3	25	0	0	1	1	0	2	0	1
C3	50	0	1	0	0	0	0	0	0
C3	1000	0	0	0	0	0	0	0	0
C4	25	0	1	2	0	1	0	0	0
C4	50	1	0	0	0	0	0	0	0
C4	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	1	0	0	1
C5	50	0	0	0	0	0	0	0	0
C5	1000	0	0	0	0	0	0	0	0

whether correct model is fitted to the data. Moreover, two-parameter models exhibit similar patterns in terms of numbers of algorithm failures, however, three-parameter models exhibit diverse patterns. As seen above, M6 and M7 have more algorithm failures in the flatter configuration C1 than those in the sharper configuration C2 for $N = 25$ and $N = 50$, while M5 and M8 exhibit the opposite. In general, almost no algorithm failure is found when sample size $N = 1000$ (there are 2 exceptions for M6 under C1) which shows asymptotic robustness of the algorithm in all models.

6.4 Asymptotic distribution for model-specific point estimator of ξ

As mentioned in §6.2, the model-specific point estimator will approach normal (Gaussian) forms, but the pattern of convergence is worth investigating. To investigate this, 8 model fits $\times$ 8 model data $\times$ 5 configurations $\times$ 3 sample-sizes $\times$ 3 estimators (AM sample mean, AM sample median or AM sample lower tercile) =

2880 histograms of $\hat{\xi}_{10}$ were constructed. (See Table 6.1 for the various configurations.) After examining all these histograms, it's found that besides the sample-size the pattern of convergence is mainly affected by the sharpness of the configured dose-response function. And in general, more normal-like distribution is observed when sample-size is larger and dose-response is sharper. When sample-size is small and the corresponding dose-response function is flat (e.g., C1 and C2), strongly right-skewed distributions usually appeared. Figure 6.2 below shows histograms of $\hat{\xi}_{10}$ using three kinds estimators when fitting M1 to its own generated data under configuration code C2 and $N = 25$.

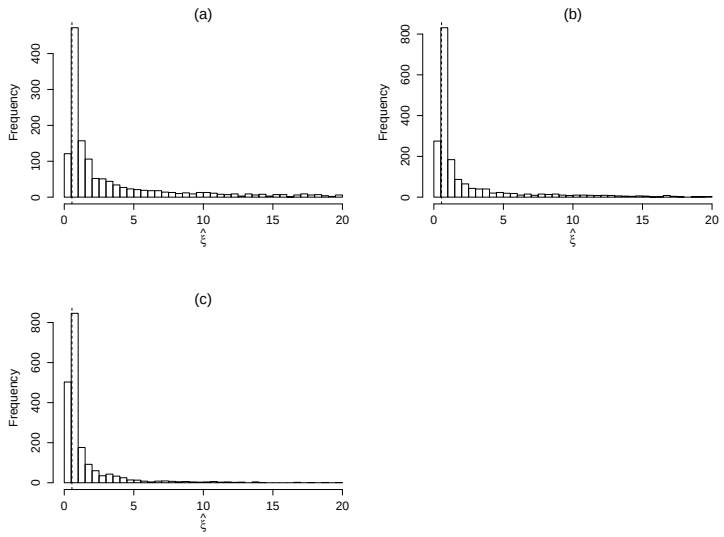

FIGURE 6.2. Histograms for various point estimates of ξ when fitting M1 to its own generated data under configuration code C2 and $N = 25$. (a) is using posterior AM sample mean; (b) is using posterior AM sample median and (c) is using posterior AM sample lower-tercile. Strongly right skewed distributions are observed. The vertical dashed line is at the true BMD. (The horizontal axis in each histogram is truncated at 20 to allow for comparable viewing. A few very large values of ξ extend beyond this range in each panel.)

It's seen that under this sample-size and configuration code, strongly right-skewed distributions are observed. Also, the AM sample lower-tercile estimate appears to be the least variable, the median estimate is slightly more variable than the lower-tercile

and the sample mean is the most variable estimator. (This is not unexpected for a right-skewed distribution that is bounded below by zero.) Therefore, the lower-tercile estimator is recommended not only because it has a better decision-theoretic understanding, but also because it is less variable.

When dose-response function is sharper, the distributions are less right-skewed and converge to normal much faster. Figure 6.3 shows the histograms of $\hat{\xi}_{10}$ using three kinds of estimators when fitting M3 to its own generated data under configuration code C3 and $N = 50$.

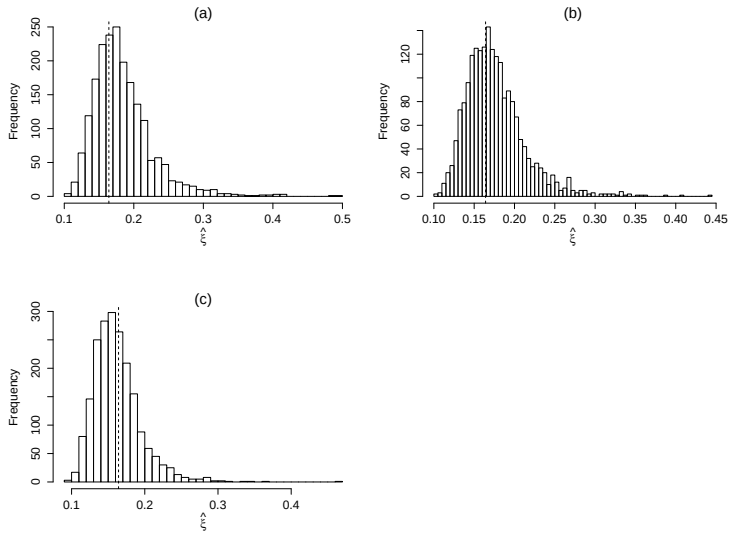

FIGURE 6.3. Histograms for various point estimates of ξ when fitting M3 to its own generated data under configuration code C3 and $N = 50$. (a) is using posterior AM sample mean; (b) is using posterior AM sample median and (c) is using posterior AM sample lower-tercile. Unimodal distributions are observed. The vertical dashed line is at the true BMD. (The horizontal axis in each histogram is truncated at 0.5 to allow for comparable viewing. A few very large values of ξ extend beyond this range in each panel.)

These distributions may still not look like normal, however, unimodality already becomes obvious at this relatively sharper configuration and variability is acceptable as well, even if N is as low as 50. In a typical toxicological experiment, each dose-level may only contain as few as 50 subjects. These histograms thus imply that in

practice, a better BMD estimate can be obtained when the estimated dose-response is relatively sharp.

When sample-size becomes quite large, say $N = 1000$, normality is usually observed and variability for these estimators becomes similar regardless of the sharpness/flatness of the dose-response function. Figure 6.4 shows the histograms of $\hat{\xi}_{10}$ using three kinds of estimators when fitting M6 to its own generated data under configuration code C2 and $N = 1000$. All three histograms look normal; the distribution for posterior mean and posterior median are roughly symmetric about the true BMD while the distribution for lower tercile is shifted slightly leftward (this is not unexpected).

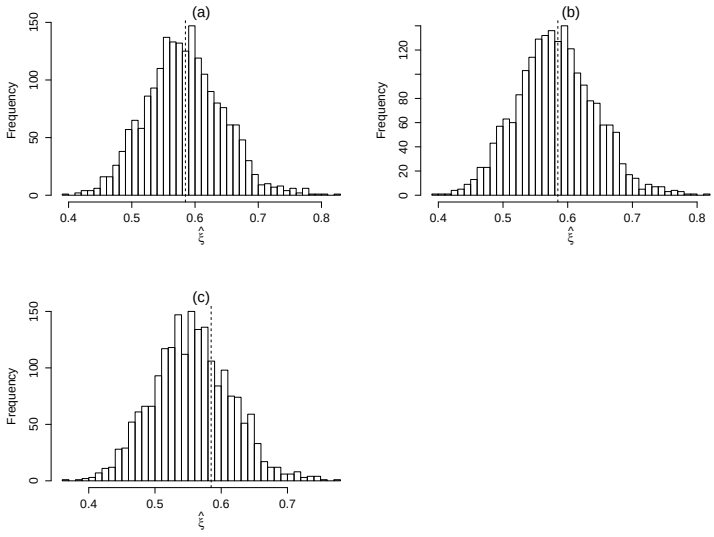

FIGURE 6.4. Histograms for various point estimates of ξ when fitting M6 to its own generated data under configuration code C2 and $N = 1000$. (a) is using posterior AM sample mean; (b) is using posterior AM sample median and (c) is using posterior AM sample lower-tercile. Normality is observed. The vertical dashed line is at the true BMD.

The above examples reflect the typical situations when a reparameterized model is fitted to the simulated data generated by itself. Regarding to converging patterns, the two-parameter models are quite similar to each other; for the three-parameter

models, M5 (two-stage model) acts similar to M8 (Weibull model) while M6 (log logistic model) acts similar to M7 (log probit model) (this kind of similarity has been ssen earlier too, when regarding to the algorithm failure patterns in §6.3), however, they all converge to normal distribution asymptotically.

To investigate the asymptotic normality on incorrect model fits, Figure 6.5 shows example histograms of $\hat{\xi}_{10}$ using three kinds of estimators when fitting M7 to data generated by M5 under configuration code C4 and $N = 1000$.

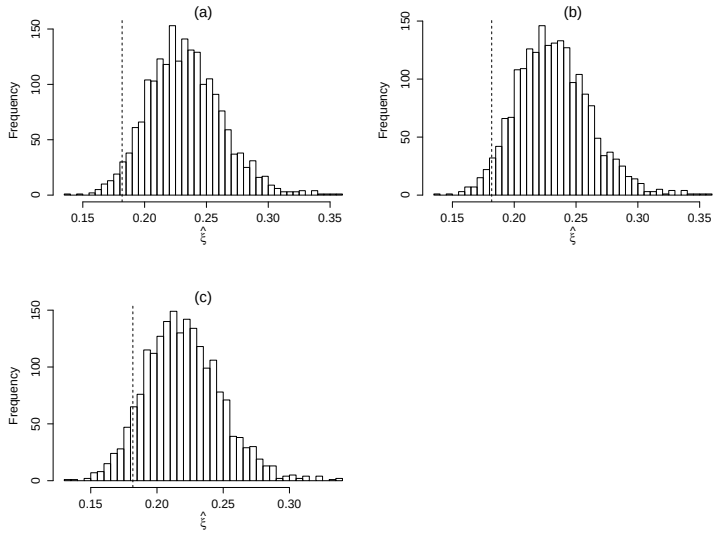

FIGURE 6.5. Histograms for various point estimates of ξ when fitting M7 to the data generated by M5 under configuration code C4 and $N = 1000$. (a) is using posterior AM sample mean; (b) is using posterior AM sample median and (c) is using posterior AM sample lower-tercile. Normality is observed, however, M7 clearly overestimate the true BMD under M5. The vertical dashed line is at the true BMD.

As seen above, the overestimation is quite obvious. In fact, when a data set is not fitted by its own model, unimodal and roughly symmetric distributions are still observed when $N = 1000$, however, the estimates may miss the true BMD badly. Indeed, in this case, asymptotic normality around the pre-specified true BMD is no longer expected.

6.5 Coverage rates for model-specific BMDL

Here, frequentist coverage was also assessed by identifying whether the generated $\underline{\xi}_{100\text{BMR}}$ was below the true value of ξ for the model under which the data were generated. The resulting empirical frequentist coverage rate is the proportion of cases that covered the true ξ out of the 2000 (some may be less than 2000, if 'algorithm failure' occurs) simulated samples. Figure 6.6 and 6.7 below display the 'spaghetti plots' of coverage rates when all 8 models were fitted to each model's generating data. The actual values for coverage rates are presented in Appendix B.

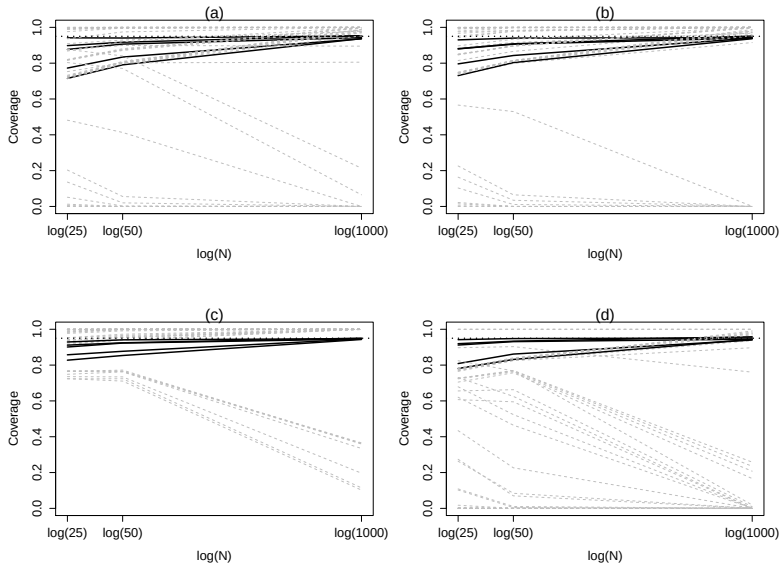

FIGURE 6.6. Spaghetti plots of empirical coverage rates for model-specific BMDL. The black lines indicate correct model fits, while the gray lines indicate incorrect model fits. Horizontal dashed-dotted line (·····) is at the nominal 95%. (a)-(d) are for data generated by M1-M4, respectively. All coverage rates under correct model fits converge to 0.95 as N increases to 1000.

The black lines monitor the changes of coverage rates when a model is correctly fitted to data generated by itself. For each model's data, there are 5 traces, each from one of the 5 configurations. Ideally, the estimated BMDL should cover the true

BMD 95% of the time, i.e., these traces should sit exactly on the 95% horizontal line, however, the coverage rates for two-parameter models at $N = 25$ or 50 are usually less than 0.95 (some are even as low as ≈ 0.8), while the coverage rates for three-parameter models are relatively higher at $N = 25$ or 50. (some are much higher than 0.95). Lower coverage rate is usually worse than higher coverage (though both are not ideal), since an estimated BMDL with actually low coverage is overly-optimistic and misleadingly-unsafe. In this sense, three-parameter models seem to perform better than two-parameter models when models are correctly fitted to the data. Asymptotically the nominal coverage rate is reached, however. These coverage rates are all very close to 0.95 at $N = 1000$ and their trends of converging to nominal are evident.

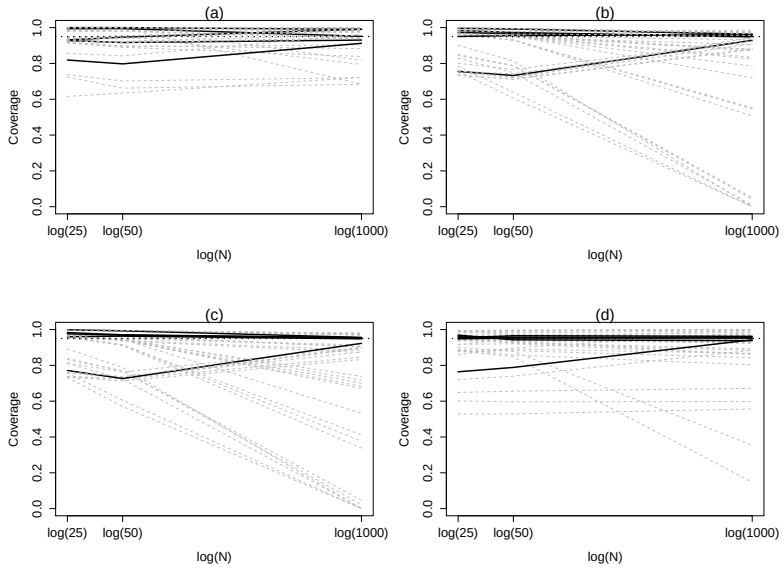

FIGURE 6.7. Spaghetti plots of empirical coverage rates for model-specific BMDL. The black lines indicate correct model fits, while the gray lines indicate incorrect model fits. Horizontal dashed-dotted line (·····) is at the nominal 95%. (a)-(d) are for data generated by M5-M8, respectively. All coverage rates under correct model fits converge to 0.95 as N increases to 1000.

The gray lines monitor the changes in coverage rates when models are incorrectly

fit to the data. There are 35 traces (5 configurations $\times$ 7 other models) for each model's data. In practice, the 'correct' model is less likely to be found to fit the data, therefore, these results are more interesting to investigate. As seen from Figures 6.6 and 6.7, the coverage rates become very unstable, some remain around 95% reference line, while more appear to be far below 95%. But in general, they perform not too bad at $N = 25$ or 50 as compared to black lines. This is not unexpected because the simulated data all look similar when N is small regardless of what model generates them and any model fit would provide similar outcome. As N increases to 1000, however, the simulated data reveals more differences between each other and some misspecified models will fit poorly. Striking results has shown that some coverage rates even drop to 0 when $N = 1000$. Along with what has been seen in Figure 6.5, it is thus shown how unreliable both BMD estimate and BMDL estimate may be when misspecified models are fitted to the data. It also shows the necessity of conquering model uncertainty in benchmark-dose analysis.

6.6 Performance of Bayesian model averaged BMDL

How the BMA BMDL compared to the corresponding generating value of ξ_{10} in Table 6.1 were studied. In some sense, it is expected that 95% of these lower credible limits to lie below ξ_{10} and using the simulations as a guide, how often this occurred was queried. Figure 6.8 displays spaghetti plots similar to those seen in §6.5 for the BMA BMDLs. There are 40 traces from 8 models and 5 configurations. Clear trend of these coverage rates converging to nominal 95% as N increases is observed. The actual values for these coverage rates are shown in Appendix B.

Considering the actual estimates of model-specific BMDLs and BMA BMDL, 40 sets of density estimators were made (8 model data $\times$ 5 configurations). Figure 6.9 provides a representative example: the figure displays modified boxplots of the 2000 simulated BMA BMDLs under model M6 (log-logistic) and configuration C3 at the

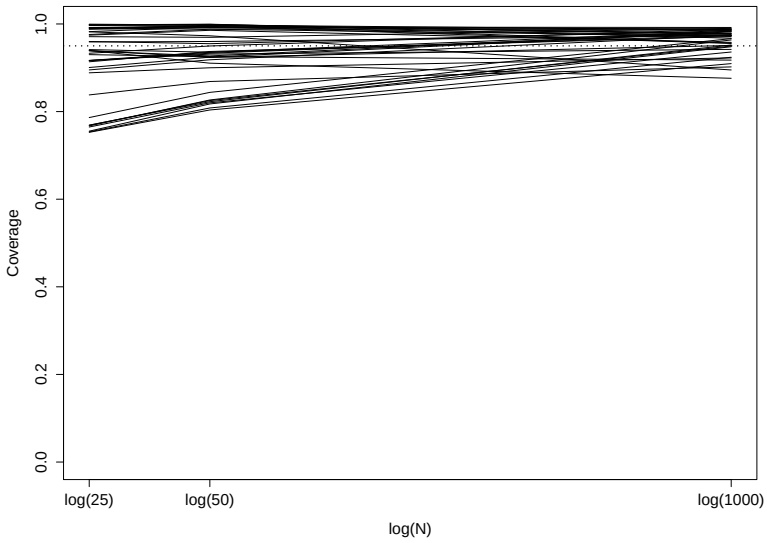

FIGURE 6.8. Spaghetti plots of empirical coverage rates for BMA BMDL. Horizontal dashed-dotted line (·····) is at the nominal 95%. All traces seem to converge to nominal asymptotically.

popular per-dose sample size of $N = 50$. Mimicking a device employed by West *et al.* (2012), the boxplots are asymmetrically 'modified' so that their upper whiskers stop at the 95^{th} percentile of the 2000 simulated BMA BMDLs from the M6/C3 combination. (The lower whiskers rest at the minimum BMDL. The hinges and median bar are the usual quartiles.) Thus the 'goal' for each boxplot is to locate its upper whisker as close to, but not exceeding, the generating value of ξ_{10}. A horizontal dashed line in the figure marks this ξ_{10} target.

In Figure 6.9 eight modified boxplots are given, corresponding to model-specific 95% BMDLs calculated under each of the eight models in Table 2.1. The additional, modified boxplot at far right gives the result for our 95% BMA BMDL. The graphic illustrates the consequences and ambiguities of single-model misspecification when calculating Bayesian BMDLs. Recall that M6 is the generating model and as anticipated, the M6 boxplot for the single-model Bayesian BMDL displays acceptable

characteristics: its modified upper whisker is slightly below the ξ_{10} target. It also exhibits a right skew, however, with median BMDL conservatively lower than any of the others in the figure.

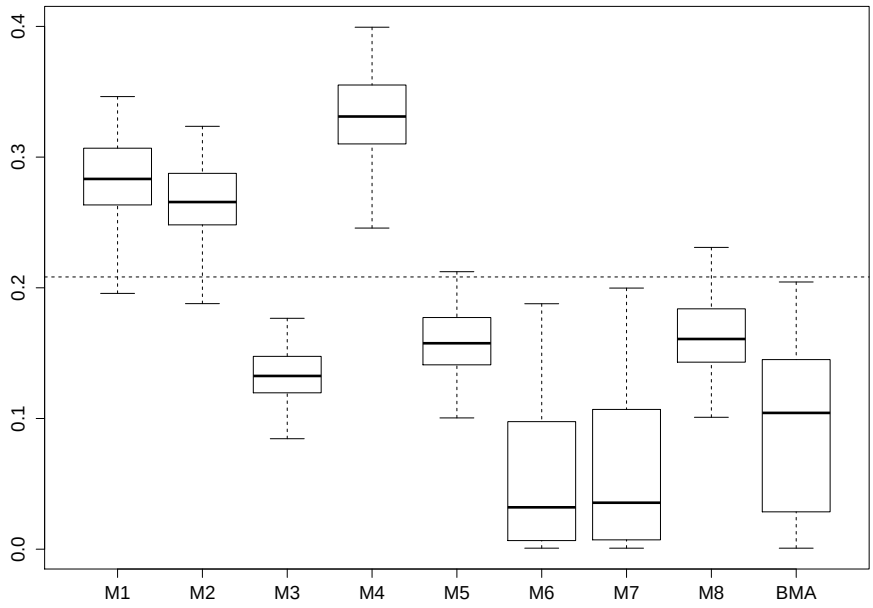

FIGURE 6.9. Modified Box plots for 95% individual-model BMDLs and BMA BMDL using simulated data from model M_6, configuration C3, and sample size $N = 50$. (See text for details.) BMR is set to 0.10. Dashed horizontal line indicates target BMD_{10} under this model configuration.

The figure also illustrates the perplexing operating characteristics of single-model BMDLs under model misspecification: when M1, M2, and (particulary) M4 are employed singly in the hierarchy, their BMDLs badly overreach the ξ_{10} target. BMDLs using models M5 and M8 are somewhat more-stable; BMDLs using model M3 show the least variation while locating conservatively below the ξ_{10} target. Analogous instances of stable/unstable single-model BMDLs occurred for all the model-configuration combinations studied, but with no clear pattern as to which models operated well or poorly when misspecified.

On balance, however, the modified boxplot for the BMA BMDL at the far right of Figure 6.9 displays reasonable operating characteristics. As desired, its modified upper whisker lies just below the target ξ_{10}, and it locates a broader collection of $\underline{\xi}_{10}$ limits closer to that target, without exceeding it, than even the correct-model M6 boxplot. (Admittedly, the M3 and perhaps M5 boxplots display even better performance here; however, these models did not exhibit consistently enhanced performance across all the configurations studied.)

It is found that unequivocal commitment to a specific dose-response model when there is any possibility of it being misspecified can lead, as often as not, to substantial overestimation of the benchmark point(s). (As described above, errors in this direction may have serious public health/safety ramifications.) Misspecifying the model can sometimes produce acceptable lower limits, but a predictable pattern was not able to be identified among these simulation results. [This corroborates similar indications by West *et al.* (2012) for single-model, frequentist BMDL calculations.] By contrast, the Bayesian BMA BMDL provides a reasonable compromise.

Performing benchmark analysis on shallow response data is difficult, when sample sizes are small – say $N = 25$ – since the pattern of response can vary greatly even if only a small deviation occurs at a single dose level. Figure 6.10 gives an example of how the Bayesian BMDLs operate in this difficult scenario. Here, the data is generated using model M8, configuration C1 (the most shallow response) and sample size of $N = 25$. In this scenario, no BMDL performs exceptionally well; indeed, for such noisy responses every chosen model and even the BMA BMDLs exhibit instability. Some small differences can be seen as M5 tends to drive BMDLs higher, M6 and M7 produce conservative estimates, and M4 and M8 perform slightly better than the others in terms of smaller IQR and higher coverage rate. Even in this situation, however, the BMA BMDL still has a roughly similar IQR compared to M4 (and smaller than M8 itself).

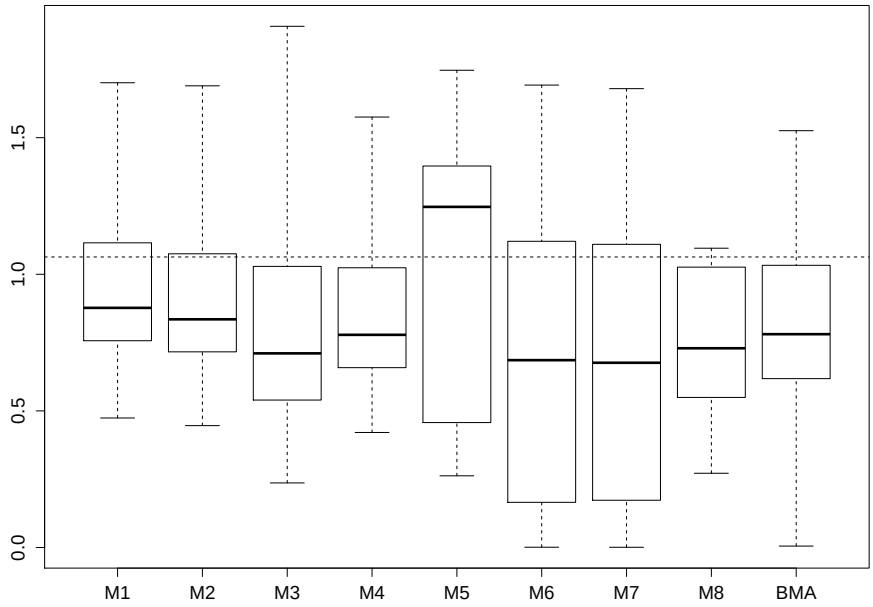

FIGURE 6.10. Modified Box plots for 95% BMDLs from individual models and Bayesian model averaging using M8 simulated data, configuration C1 and sample size 25. (See text for details of the modifications.) The dashed line represents the true value of BMD under this setting. In this scenario, no individual BMDL performs exceptionally well with M4 and M8 perform slightly better than the others in terms of smaller IQR and higher coverage, however BMA BMDL still has a roughly similar IQR compared to M4 and smaller than M8.

Figure 6.11 extends this analysis to the large-sample setting, displaying a similar graphic when the per-dose sample size is $N = 1000$. Here, results for model M3 under configuration C5 are highlighted. As what might be expected with such a large N, no single-model fit exhibits acceptable performance, save that of the correct model (M3) and possibly models M5 and M8. (Notice that model M3 is a special case of both M5, with $\beta_2 = 0$, and M8, with $\beta_1 = 1$, in Table 2.1. As in Figure 6.9, these three models sometimes perform similarly, although it was also found that cases where their operating characteristics diverged.) Moving to the BMA BMDL, however, overcomes

these negative characteristics: the BMA BMDL boxplot at far right is almost identical to that for the correct-model M3 boxplot. Here again, the BMA BMDL provides valuable robustness in the presence of model uncertainty/misspecification.

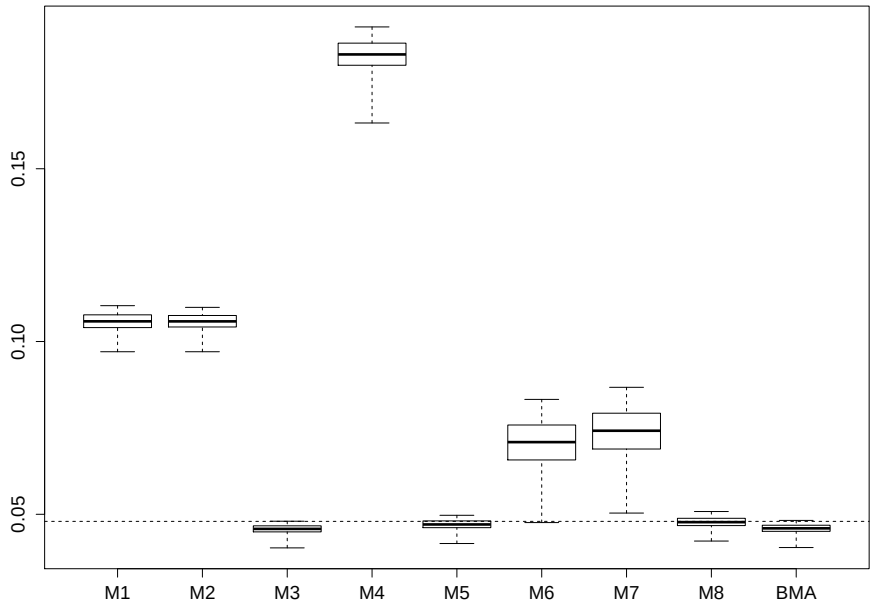

FIGURE 6.11. Modified Box plots for 95% individual-model BMDLs and BMA BMDL using simulated data from model M_3, configuration C5, and sample size $N = 1000$. (See text for details.) BMR is set to 0.10. Dashed horizontal line indicates target BMD_{10} under this model configuration.

Chapter 7

Summary and Conclusion

The final chapter of this dissertation serves as a summary of the work presented in the previous chapters.

In the first chapter of the dissertation, an overview of the process of quantitative risk assessment and the steps it entails was given. The dose-response modeling and quantitative risk estimation was addressed. Here, risk $R(d)$ was defined as the probability of a subject exhibiting a pre-defined adverse effect when exposed to a particular dose level, d, of a hazardous agent. Two key assumptions were made on this fundamental definition of risk. The first assumption was that risk is a monotone increasing function. The second assumption was that non-zero background risk at $d = 0$ may exist in the population of interest. To correct for this spontaneous background risk, excess risks such as additional risk and extra risk were further introduced. These excess risks are more commonly used in quantitative risk estimation. After that, difficulties in low-dose risk estimation were discussed and a modern technique—benchmark dose analysis—was introduced. In benchmark dose analysis, one uses the functional specification for $R(d)$ to provide low-dose estimates for risk and/or excess risk. To estimate the benchmark dose (BMD) and the corresponding benchmark dose lower limit (BMDL) in quantal-response data setting (as is the focus of this dissertation), the traditional frequentist method through parametric maximum likelihood estimation was first introduced, and then a thorough literature review on established parametric hierarchical Bayesian benchmark dose analysis was provided.

Chapter 2 introduced 8 popular quantal-response models as seen in the U.S. EPA's BMDS software (Davis *et al.*, 2012). These were forms of generalized linear models expressed via traditional $\beta-$type parameters. Because these parameters usually don't

have pertinent risk analytic interpretations, it is difficult to construct prior distributions for them. As a result, objective, even improper, priors had to be used instead. In order to embrace useful prior information into the Bayesian hierarchical modeling, reparameterizations for these 8 models were performed by using the target quantity, BMD (ξ), background risk (γ_0) and in three-parameter models, also a risk at some dose level (usually the risk at the highest level, γ_1) as the new parameters. Mathematical forms have been derived; these reparameterizations present more burdensome notation for $R(d)$, however, allow benchmark analysts to formulate a clearer and more application-oriented hierarchical model, from which to produce inferences on BMD.

The hierarchical Bayesian modeling framework was introduced in Chapter 3. This included three major topics: prior specification, posterior approximation and Bayesian estimation. For positive quantity ξ, inverse gamma prior was specified by default; for probability quantities γ_0 and γ_1 (also for the risks at any other dose level), Beta prior was specified. Prior elicitation process using first quartile and median was introduced. When elicitation is not available, proper objective priors using $\xi \sim IG(0.001, 0.001)$ and $\gamma_0, \gamma_1 \sim Beta(\frac{1}{2}, \frac{1}{2})$ were employed and recommended. Joint posterior density was approximated by using an adaptive Metropolis algorithm introduced in Andrieu and Thoms (2008), convergence diagnosis and burn-in determination were applied mimicking a method introduced in Geweke (1992). Several decision theoretical Bayesian estimators were discussed and the posterior lower tercile estimator resulted from bi-linear asymmetric loss function was recommended for BMD point estimation. With the AM sample, 95% BMDL was easily estimated as the 5th lower percentile of the AM sample. To investigate and assess the prior sensitivity on ξ, an $\epsilon-$contamination study followed by O'Hagan (1994, §7.15) was introduced.

The increasing number of available models for benchmark dose analysis brings in the issue of model uncertainty and adequacy. Chapter 4 discussed this important issue. Due to the difficulty of finding reliable selecting criterion for model selection, a Bayesian model averaging (BMA) method introduced by Hoeting *et al.* (1999)

was employed. In this method, those 8 quantal-response models were chosen as the uncertainty class and a mixture distribution for ξ was constructed using the posterior model probabilities as the weights. The BMA BMD is not pursued while the BMA BMDL which satisfies the corresponding probability statement is approximated by using a 'direct method' from each model-specific AM sample.

In Chapter 5, application of the proposed hierarchical Bayesian method was illustrated via a real carcinogenicity study performed by U.S. National Toxicology Program (NTP) on the chemical cumene. BMD and BMDL were first estimated by using the popular quantal linear model. To investigate prior sensitivity, three $\epsilon-$contaminating scenarios were considered, and it was shown that the estimates from quantal linear model was reasonably robust. After that, another popular model—logistic model—was fitted to the data and its estimates was found to be significantly different from quantal linear model's. Facing the issue of model uncertainty, the remaining 6 models were then fitted to the data and the Bayesian model averaged BMDL was produced. This showed an example that BMA adjustment could free the risk assessors from model inadequacies and inferential uncertainties which were frequently encountered when committing to a single-model benchmark analysis.

Chapter 6 aimed to evaluate the performance of the proposed method via a simulation study. 8 reparameterized quantal response models were fitted to a set of 2000 data sets generated by each of these models under 5 configurations and 3 sample-sizes. Asymptotic normality of the point estimates of BMD appeared to be satisfied when fitting the correct model to the data. The model-specific BMDL also appeared to converged to nominal as N increased when the correct model was fitted to the data. When model were misspecified, however, both the point estimates and the BMDL became unreliable. The BMA BMDL, on the other hand, could provide much more stable and acceptable estimates.

With the conclusion of this portion of the research, there are several natural extensions for further study. Mimicking previous Bayesian benchmark analysis methods

(Shao and Small, 2011, 2012; Shao, 2012), the current method assumed that the unknown parameters enter the joint prior independently. The correlation between parameters may not be ignored, however. If correlation between parameters can be elicited and successfully incorporated into the prior hierarchy, the BMD/BMDL estimation may be further improved. In addition, objective prior specification can be approached via many strategies when elicitation breaks down. The current method assumed independent $IG(0.001, 0.001)$ and $Beta(\frac{1}{2}, \frac{1}{2})$ priors. Other forms may be pertinent, however, and these choices might still bring possible subjectivity to the proposed hierarchical model. It is also of interest to investigate how the proposed approach operates under different design configurations. The current research focused on a geometric, four-dose design, arguably the quintessential standard in cancer and laboratory-animal toxicology testing. Greater information might be gained about the pattern of dose response, however, and therefore about the BMD, if the number of doses is increased and/or the dose spacings are changed. Experimental design for dose-response studies with focus on the BMD is an emerging area in the statistical literature (Muri *et al.*, 2009; Öberg, 2010; Sand *et al.*, 2008; Shao and Small, 2012) and how to optimally design/allocate experimental resources for BMD estimation and inferences under a Bayesian paradigm is an open question.

Appendix A

Numbers of algorithm failures under remaining model fits

Chapter 6 listed the numbers of data failures for data generated from each model and the numbers of algorithm failures when fitting M1, M5 and M7 to the simulated data sets. Appendix A completes those tables by adding the numbers of algorithm failures for the remaining model-fits (M2-M4, M6 and M8). Model codes are from Table 2.1.

Table A.1. Numbers of algorithm failures when fitting M2 to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	30	24	16	33	27	31	24	25
	50	12	12	13	10	8	9	11	13
	1000	0	0	0	0	0	0	0	0
C2	25	15	8	16	11	17	14	19	18
	50	9	8	9	5	3	4	5	4
	1000	0	0	0	0	0	0	0	0
C3	25	0	0	0	0	0	2	1	2
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	3	3	1	2	4	2	2	1
	50	1	0	0	0	0	1	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

TABLE A.2. Numbers of algorithm failures when fitting M3 to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	18	13	19	12	15	12	10	13
	50	6	5	6	6	7	9	6	2
	1000	0	0	0	0	0	0	0	0
C2	25	6	5	15	9	16	15	10	12
	50	2	7	10	3	7	4	5	3
	1000	0	0	0	0	0	0	0	0
C3	25	1	1	3	0	0	0	1	1
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	3	3	0	3	3	0	1	1
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	1	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

TABLE A.3. Numbers of algorithm failures when fitting M4 to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	27	21	18	26	29	14	36	24
	50	18	16	16	9	19	11	11	9
	1000	0	0	0	0	0	0	0	0
C2	25	25	20	15	27	16	19	19	16
	50	6	13	11	5	10	8	8	11
	1000	0	0	0	0	0	0	0	0
C3	25	0	0	2	0	3	3	2	2
	50	0	1	0	0	0	1	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	4	5	13	1	3	7	5	3
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

TABLE A.4. Numbers of algorithm failures when fitting M6 to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	83	84	110	114	94	106	128	114
	50	50	46	82	54	53	50	47	58
	1000	0	1	0	0	0	1	0	0
C2	25	32	37	35	41	39	39	46	38
	50	37	32	26	25	29	23	35	31
	1000	0	0	0	0	0	0	0	0
C3	25	2	1	0	1	0	0	1	0
	50	0	0	1	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C4	25	0	1	1	1	0	0	0	1
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0
C5	25	1	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

TABLE A.5. Numbers of algorithm failures when fitting M8 to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	6	6	15	6	4	5	3	6
	50	2	4	8	1	2	1	0	1
	1000	0	0	0	0	0	0	0	0
C2	25	59	57	73	52	62	59	65	63
	50	31	22	48	20	18	15	20	28
	1000	0	0	0	0	0	0	0	0
C3	25	0	1	2	0	1	1	4	0
	50	0	0	0	0	0	1	0	1
	1000	0	0	0	0	0	0	0	0
C4	25	1	0	3	1	0	0	0	1
	50	0	0	0	0	0	0	0	1
	1000	0	0	0	0	0	0	0	0
C5	25	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0

Appendix B

Coverage rates for model specific BMDL and BMA BMDL

Chapter 6 illustrated the coverage rates for model specific BMDL and BMA BMDL via spaghetti plots (See Figures 6.6, 6.7 and 6.8). Here, the actual values of coverage rates are provided supplement to those plots. Model codes are from Table 2.1.

Table B.1. Coverage rates when fitting logistic model (M1) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.715	0.714	0.732	0.726	0.723	0.722	0.726	0.726
	50	0.791	0.797	0.807	0.801	0.798	0.810	0.803	0.801
	1000	0.937	0.953	0.989	0.969	0.973	0.979	0.987	0.978
C2	25	0.773	0.753	0.482	0.868	0.820	0.817	0.817	0.820
	50	0.834	0.802	0.413	0.920	0.881	0.875	0.871	0.876
	1000	0.936	0.805	0.000	1.000	0.998	0.996	0.985	0.997
C3	25	0.898	0.803	0.002	0.909	0.004	0.052	0.136	0.012
	50	0.917	0.770	0.000	0.929	0.000	0.002	0.020	0.000
	1000	0.954	0.065	0.000	0.979	0.000	0.000	0.000	0.000
C4	25	0.877	0.873	0.203	0.999	0.936	0.990	0.998	0.976
	50	0.906	0.899	0.056	1.000	0.975	1.000	1.000	0.996
	1000	0.942	0.895	0.000	1.000	1.000	1.000	1.000	1.000
C5	25	0.942	0.876	0.000	0.996	0.989	1.000	1.000	0.994
	50	0.940	0.841	0.000	0.999	0.994	1.000	1.000	0.998
	1000	0.952	0.213	0.000	1.000	1.000	1.000	1.000	1.000

TABLE B.2. Coverage rates when fitting probit model (M2) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.733	0.731	0.750	0.740	0.736	0.746	0.746	0.746
	50	0.799	0.803	0.815	0.811	0.811	0.812	0.815	0.815
	1000	0.915	0.936	0.988	0.953	0.962	0.973	0.981	0.972
C2	25	0.815	0.796	0.566	0.897	0.847	0.846	0.849	0.851
	50	0.866	0.843	0.530	0.933	0.899	0.896	0.895	0.895
	1000	0.979	0.940	0.000	1.000	1.000	0.998	0.996	0.999
C3	25	0.951	0.880	0.004	0.967	0.014	0.103	0.226	0.022
	50	0.979	0.908	0.000	0.983	0.000	0.011	0.065	0.000
	1000	1.000	0.952	0.000	1.000	0.000	0.000	0.000	0.000
C4	25	0.885	0.880	0.164	0.998	0.942	0.991	0.997	0.979
	50	0.916	0.906	0.034	1.000	0.979	1.000	1.000	0.997
	1000	0.968	0.940	0.000	1.000	1.000	1.000	1.000	1.000
C5	25	0.975	0.930	0.000	0.999	0.995	1.000	1.000	0.998
	50	0.983	0.940	0.000	1.000	0.999	1.000	1.000	0.999
	1000	1.000	0.942	0.000	1.000	1.000	1.000	1.000	1.000

TABLE B.3. Coverage rates when fitting quantal linear model (M3) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.724	0.736	0.827	0.723	0.769	0.764	0.749	0.764
	50	0.711	0.734	0.854	0.724	0.766	0.772	0.762	0.765
	1000	0.102	0.197	0.943	0.114	0.365	0.361	0.335	0.359
C2	25	0.923	0.915	0.858	0.952	0.942	0.939	0.940	0.939
	50	0.940	0.941	0.877	0.966	0.956	0.954	0.952	0.955
	1000	1.000	1.000	0.946	1.000	1.000	1.000	1.000	1.000
C3	25	0.999	0.998	0.911	0.999	0.940	0.977	0.989	0.953
	50	1.000	1.000	0.924	1.000	0.961	0.991	0.996	0.973
	1000	1.000	1.000	0.951	1.000	0.998	1.000	1.000	1.000
C4	25	0.983	0.983	0.901	0.999	0.989	0.998	0.999	0.995
	50	0.999	0.999	0.923	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	0.946	1.000	1.000	1.000	1.000	1.000
C5	25	1.000	1.000	0.929	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	0.941	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	0.948	1.000	1.000	1.000	1.000	1.000

TABLE B.4. Coverage rates when fitting quantal quadratic model (M4) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.765	0.768	0.772	0.781	0.776	0.782	0.783	0.783
	50	0.825	0.827	0.828	0.831	0.831	0.840	0.839	0.837
	1000	0.897	0.936	0.991	0.944	0.969	0.978	0.984	0.975
C2	25	0.656	0.607	0.263	0.809	0.726	0.723	0.705	0.726
	50	0.663	0.595	0.084	0.861	0.770	0.763	0.757	0.768
	1000	0.009	0.000	0.000	0.940	0.258	0.207	0.167	0.235
C3	25	0.897	0.731	0.000	0.919	0.000	0.000	0.002	0.000
	50	0.903	0.625	0.000	0.933	0.000	0.000	0.000	0.000
	1000	0.760	0.000	0.000	0.958	0.000	0.000	0.000	0.000
C4	25	0.003	0.002	0.000	0.912	0.018	0.434	0.621	0.109
	50	0.000	0.000	0.000	0.931	0.000	0.227	0.465	0.010
	1000	0.000	0.000	0.000	0.945	0.000	0.000	0.000	0.000
C5	25	0.274	0.102	0.000	0.942	0.679	1.000	1.000	0.825
	50	0.070	0.004	0.000	0.948	0.523	1.000	1.000	0.767
	1000	0.000	0.000	0.000	0.952	0.000	1.000	1.000	0.019

TABLE B.5. Coverage rates when fitting two-stage model (M5) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.914	0.918	0.941	0.918	0.928	0.927	0.925	0.927
	50	0.890	0.898	0.954	0.894	0.917	0.919	0.923	0.916
	1000	0.839	0.883	0.986	0.904	0.930	0.942	0.959	0.938
C2	25	0.932	0.928	0.950	0.924	0.931	0.933	0.942	0.927
	50	0.949	0.950	0.953	0.945	0.947	0.952	0.945	0.951
	1000	0.987	0.980	0.794	1.000	0.991	0.987	0.977	0.988
C3	25	1.000	0.999	0.738	1.000	0.819	0.939	0.979	0.857
	50	1.000	1.000	0.702	1.000	0.797	0.938	0.981	0.842
	1000	1.000	0.974	0.721	1.000	0.912	0.993	1.000	0.962
C4	25	0.983	0.982	0.725	0.999	0.994	1.000	1.000	0.999
	50	0.982	0.980	0.662	1.000	0.995	1.000	1.000	1.000
	1000	0.948	0.941	0.683	1.000	0.952	1.000	1.000	0.988
C5	25	1.000	1.000	0.616	1.000	1.000	1.000	1.000	1.000
	50	1.000	0.999	0.634	1.000	1.000	1.000	1.000	1.000
	1000	0.818	0.684	0.722	1.000	0.990	1.000	1.000	0.999

TABLE B.6. Coverage rates when fitting log-logistic model (M6) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.738	0.732	0.801	0.751	0.754	0.756	0.759	0.757
	50	0.711	0.713	0.767	0.725	0.722	0.733	0.721	0.742
	1000	0.878	0.882	0.922	0.927	0.909	0.929	0.946	0.926
C2	25	0.996	0.997	0.995	0.998	0.998	0.996	0.999	0.998
	50	0.995	0.992	0.989	0.994	0.995	0.992	0.994	0.995
	1000	0.975	0.975	0.970	0.960	0.971	0.963	0.940	0.966
C3	25	0.969	0.966	0.973	0.973	0.969	0.973	0.981	0.967
	50	0.947	0.948	0.958	0.954	0.955	0.966	0.977	0.957
	1000	0.905	0.879	0.870	0.931	0.835	0.954	0.980	0.878
C4	25	0.964	0.961	0.968	0.980	0.965	0.984	0.985	0.971
	50	0.930	0.930	0.954	0.951	0.930	0.973	0.979	0.951
	1000	0.545	0.507	0.720	0.825	0.553	0.948	0.986	0.785
C5	25	0.785	0.754	0.901	0.850	0.829	0.951	0.961	0.845
	50	0.639	0.607	0.814	0.788	0.749	0.956	0.964	0.786
	1000	0.000	0.000	0.000	0.051	0.008	0.962	0.957	0.042

TABLE B.7. Coverage rates when fitting log-probit model (M7) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.735	0.739	0.814	0.738	0.756	0.764	0.771	0.765
	50	0.713	0.724	0.763	0.724	0.732	0.732	0.727	0.737
	1000	0.833	0.846	0.900	0.896	0.875	0.897	0.923	0.895
C2	25	0.999	0.998	0.995	0.998	0.997	0.998	0.998	0.997
	50	0.994	0.992	0.989	0.994	0.995	0.994	0.993	0.995
	1000	0.978	0.981	0.974	0.970	0.977	0.970	0.956	0.971
C3	25	0.964	0.962	0.964	0.969	0.961	0.967	0.976	0.962
	50	0.940	0.936	0.949	0.948	0.942	0.960	0.969	0.942
	1000	0.884	0.872	0.718	0.914	0.683	0.902	0.954	0.739
C4	25	0.957	0.957	0.965	0.973	0.959	0.981	0.983	0.968
	50	0.914	0.910	0.945	0.940	0.915	0.963	0.971	0.942
	1000	0.376	0.338	0.532	0.701	0.411	0.904	0.954	0.671
C5	25	0.757	0.733	0.891	0.839	0.811	0.949	0.960	0.832
	50	0.606	0.570	0.788	0.769	0.723	0.949	0.963	0.768
	1000	0.000	0.000	0.000	0.046	0.004	0.954	0.948	0.022

TABLE B.8. Coverage rates when fitting Weibull model (M8) model to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.949	0.960	0.972	0.965	0.967	0.968	0.965	0.969
	50	0.924	0.928	0.966	0.932	0.942	0.948	0.949	0.943
	1000	0.884	0.892	0.985	0.934	0.927	0.942	0.957	0.938
C2	25	0.944	0.935	0.889	0.962	0.951	0.949	0.956	0.953
	50	0.956	0.941	0.858	0.973	0.966	0.965	0.959	0.965
	1000	0.973	0.967	0.803	0.954	0.967	0.957	0.930	0.962
C3	25	0.963	0.953	0.649	0.966	0.720	0.860	0.914	0.764
	50	0.954	0.945	0.656	0.958	0.740	0.896	0.951	0.788
	1000	0.940	0.922	0.671	0.952	0.885	0.984	0.996	0.942
C4	25	0.882	0.876	0.601	0.989	0.918	0.985	0.991	0.959
	50	0.891	0.882	0.597	0.972	0.920	0.982	0.990	0.957
	1000	0.868	0.843	0.598	0.956	0.862	0.994	0.999	0.953
C5	25	0.899	0.880	0.528	0.950	0.938	0.994	0.993	0.948
	50	0.878	0.848	0.529	0.949	0.933	0.997	0.997	0.952
	1000	0.354	0.147	0.556	0.954	0.863	1.000	1.000	0.954

TABLE B.9. Coverage rates when applying Bayesian model averaging to the simulated data generated by 8 models, under 5 configurations and 3 sample-sizes.

Configuration	N	M1	M2	M3	M4	M5	M6	M7	M8
C1	25	0.753	0.754	0.786	0.755	0.765	0.769	0.767	0.769
	50	0.804	0.808	0.844	0.817	0.820	0.823	0.827	0.825
	1000	0.909	0.924	0.967	0.948	0.936	0.952	0.964	0.950
C2	25	0.901	0.895	0.838	0.933	0.916	0.914	0.917	0.917
	50	0.924	0.918	0.869	0.950	0.933	0.937	0.931	0.935
	1000	0.982	0.974	0.902	0.988	0.979	0.972	0.942	0.974
C3	25	0.990	0.982	0.940	0.991	0.938	0.960	0.971	0.942
	50	0.989	0.973	0.923	0.993	0.927	0.955	0.970	0.935
	1000	0.989	0.895	0.951	0.992	0.984	0.985	0.990	0.984
C4	25	0.976	0.975	0.930	0.998	0.983	0.997	1.000	0.989
	50	0.988	0.986	0.925	0.997	0.993	0.999	0.998	0.997
	1000	0.973	0.958	0.917	0.979	0.985	0.976	0.987	0.978
C5	25	0.959	0.940	0.888	0.992	0.988	0.998	0.998	0.992
	50	0.960	0.910	0.900	0.994	0.992	1.000	0.999	0.995
	1000	0.974	0.876	0.922	0.988	0.974	0.968	0.957	0.956

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory,* Petrov BN, Csaki B (eds). Akademiai Kiado: Budapest; 267–281.

Ando, Tomohiro (2010). *Bayesian Model Selection and Statistical Modeling*, 1st edn. Boca Raton, FL: Chpman & Hall/CRC

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing* **18**, 343–383.

Armitage, P. and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* **8**, 1–12.

Barnes, K. (2012), Markov chain Monte Carlo (MCMC). In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics, vol. 3, 2nd Edn.*, 1524–1532. Chichester: John Wiley & Sons.

Barzilai, J. and Borwein, J. M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis* **8**, 141–148.

Bailer, A. J., Noble, R. B. and Wheeler, M. W. (2005). Model uncertainty and risk estimation for experimental studies of quantal responses. *Risk Analysis* **25**, 291–299.

Bailer, A. J. and Portier, C. J. (1994). Modeling risks from water contaminants: The application of concentration-response models. In G. Rhoda and M. Wang (eds), *Water Contamination and Health: Integration of Exposure Assessment*, 447–466, New York: M. Dekker.

Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Annals of Statistics* **37**, 2011–2055.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society, series B (Methodological)* **41**, 113–147.

Brown, H. and Prescott, R. (2006). *Applied Mixed Models in Medicine*, Chichester: John Wiley & Sons.

Boruff, B. J., Easoz, J. A., Jones, S. D., Landry, H. R., Mitchem, J. D., and Cutter, S. L. (2003). Tonado hazards in the United States. *Climate Research*, **24**, 103–117.

Brown, C. and Koziol, J. (1983). Statistical aspects of the estimation of human risk from suspected environmental carcinogens, *SIAM Review*, **25**, 151–181.

Buckley, B. E., Piegorsch, W. W. and West, R. W. (2009). Confidence limits on one-stage model parameters in benchmark risk assessment. *Environmental and Ecological Statistics* **16**, 53–62.

California Office of Environmental Health Hazard Assessment (2004). *Proposition 65 Maximum Allowable Dose Level (MADL) for Reproductive Toxicity for Methyl Bromide as a Structural Fumigant*. Sacramento, CA: CA Office of Environmental Health Hazard Assessment (OEHHA), Reproductive and Cancer Hazard Assessment Section.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd edn. Pacific Grove, CA: Duxbury.

Chao, M. T. (1970). The Asymptotic Behavior of Bayes' Estimators. *The Annals of Mathematical Statistics* **41**, 601–608.

Chen, D. G. (2010). Incorporating historical control information into quantal bioassay with Bayesian approach. *Computational Statistics & Data Analysis* **54**, 1646–1656.

Chen, J. J., Kodell, R. L. (1989). Quantitative risk assessment for teratological effects, *Journal of the American Statistical Association*, **84**, 966-971.

Christensen, R., Johnson, W. O., Branscum, A. J. and Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton, FL: Chapman & Hall/CRC Press.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comprarative review. *Journal of the American Statistical Association*, **91**, 883–904.

Crump, K. S., Hoel, D. G., Langley, C. H. and Peto, R. (1976). Fundamental carcinogenic processes and their implications for low dose risk assessment, *Cancer Research*, **36**, 2973–2979.

Crump, K. S. and Howe, R. (1985). A review of methods for calculating confidence limits in low dose extrapolation. In Clayson, D. B., Krewski, D. and Munro, I. (eds.), *Toxicological Risk Assessment, Volume I: Biological and Statistical Criteria*, 187–203. Boca Raton, FL: CRC Press.

Davis, J. A., Gift, J. S. and Zhao, Q. J. (2012). Introduction to benchmark dose methods and U.S. EPA's Benchmark Dose Software (BMDS) version 2.1.1. *Toxicology and Applied Pharmacology* **254**, 181–191.

European Union (2003). *Technical Guidance Document (TGD) on Risk Assessment of Chemical Substances following European Regulations and Directives, Parts I-IV*. Technical Report number EUR 20418 EN/1-4. Ispra, Italy: European Chemicals Bureau (ECB).

Faes, C., Geyes, H., Aerts, M. and Molenberghs, G. (2006). A hierarchical modeling approach for risk assessment in developmental toxicity studies, *Computational Statistics and Data Analysis*, **51**, 1848–1861.

Faustman, E. M. and Bartell, S. M. (1997). Review of noncancer risk assessment: Applications of benchmark dose methods. *Human and Ecological Risk Assessment* **3**, 893–920.

Flegal, J. M. and Jones, G., L. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, **38** 1034–1070.

Foronda, N. M., Fowles, J., Smith, N., Taylor, M. and Temple, W. (2007). A benchmark dose analysis for sodium monofluoroacetate (1080) using dichotomous toxicity data. *Regulatory Toxicology and Pharmacology* **47**, 84–89.

Fox, D. R. (2006). Statistical issues in ecological risk assessment, em Human and Ecological Risk Assessment, **12**, 120–129.

Gaylor, D. W. (1989). Quantitative risk analysis for quantal reproductive and developmental effects, *Environmental Health Perspectives*, **79**, 243–246.

Gaylor, D. W. (1998). Dose-response models in risk analysis. In Armitage, P. and Colton, T. (eds.), *Encyclopedia of Biostatistics* **2**, 1219–1223. Chichester: John Wiley & Sons

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association* **95**, 1300–1304.

Gelman, A., Carlin, B. J., Stein, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edn. Boca Raton, FL: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics* **4**, 169–193. Oxford: Oxford University Press.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, Hoboken: John Wiley & Sons.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics).* **41**, 337–348.

Grieve, A. P. (1988). A Bayesian approch to the analysis of LD50 experiments. In Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M. (eds.), *Baysian Statistics* **3**, 617–630. Oxford: Oxford University Press.

Grimston, M. (2002). Nuclear risk. In A. H. El-Shaarawi and W. W. Piegorsch (eds), *Encyclopedia of Environmetrics*, Chichester: John Wiley & Sons, **3**, 1425-1432.

Guha, N., Roy, A., Kopylev, L., Fox, J., Spassova, N., and White, P. (2013). Nonparametric Bayesian methods for benchmark dose estimation. *Risk Analysis* **9**, 1608–1619.

Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli.* **7** 223–242.

Haseman, J. K. (1984). Statistical issues in the design, analysis and interpretation of animal carcinogenicity studies, *Environmental Health Perspectives*, **58**, 385–392.

Haseman, J. K. (1985). Issues in carcinogenicity testing: Dose selection, *Fundamental and Applied Toxicology*, **5**, 66–78.

Heidelberger, P and Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* **24**, 233–245.

Held, L. (2004). Simultaneous inference in risk assessment; a Bayesian perspective. In Antoch, J. (ed.), *COMPSTAT 2004. Proceedings in Computational Statistics*, 213–222. Heidelberg: Physica-Verlag.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science* **14** 382–401. (corr. **15** 193–195).

Hu, B., Ji, Y. and Tsui, K.-W. (2008). Bayesian estimation of inverse dose response. *Biometrics* **64**, 1223–1230.

Izadi, H., Grundy, J. E. and Bose, R. (2012). Evaluation of the benchmark dose for point of departure determination for a variety of chemical classes in applied regulatory settings. *Risk Analysis* **32**, 830–835.

Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford: Oxford University Press.

Kang, S. H., Kodell. R. L. and Chen, J. J. (2000). Incorporating model uncertainties along with data uncertainties in microbial risk assessment. *Regulatory Toxicology and Pharmacology* **32**, 68–72.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factor. *Journal of the American Statistics Association* **90**, 773–795.

Kodell, R. L. (2005). Managing uncertainty in health risk assessment. *International Journal of Risk Assessment and Management* **14**, 193–205.

Krewski, D., Gaylor D. and Szyszkowicz, M. (1991). A model-free approach to low dose extrapolation, *Environmental Health Perspectives*, **90**, 279-285.

Krewski, D., van Ryzin, J. (1981). Risk assessment, quantitative. In M. Csörgö, D. A. Dawson, J. N. K Rao and A. K. M. E. Saleh (eds), *Statistics and Related Topics*, 201–231, Chichester: John Wiley & Sons.

Kuhnert, P. M. (2011). Four case studies in using expert opinion to inform priors. *Environmetrics* **22**, 662–674.

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

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* **24**, 2401–2428.

Li, J., Zhang, C., Nordheim, E. V. and Lehner, C. E. (2008). On the multivariate predictive distribution of multi-dimensional effective dose: a Bayesian approach. *Journal of Statistical Computation and Simulation* **87**, 429–442.

Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67.

Meng, X. and Wong W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica* **6**, 831–860.

Moerbeek, M., Piersma, A. H. and Slob, W. (2004). A comparison of three methods for calculating confidence intervals for the benchmark dose. *Risk Analysis* **24**, 31–40.

Morales, K. H., Ibrahim, J. G., Chen, C.-J. and Ryan, L. M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association* **101**, 9–17.

Muri, S. D., Schlatter, J. R. and Brüschweiler, B. J. (2009). The benchmark dose approach in food risk assessment: Is it applicable and worthwhile? *Food and Chemical Toxicology* **47**, 2906–2925.

Naufal, Z., Kathman, S. and Wilson, C. (2009). Bayesian derivation of an oral cancer slope factor distribution for 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Regulatory Toxicology and Pharmacology* **55**, 69–75.

Öberg M. (2010). Benchmark dose approaches in chemical health risk assessment in relation to number and distress of laboratory animals. *Regulatory Toxicology and Pharmacology* **58**, 451–454.

OECD (2006). *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application, Series on Testing and Assessment No. 54.* Paris: Environment Directorate, Organisation For Economic Co-Operation and Development.

OECD (2008). *Draft Guidance Document on the Performance of Chronic Toxicity and Carcinogenicity Studies, Supporting TG 451, 452 and 453.* Paris: Organisation For Economic Co-Operation and Development.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics, Volume 2B, Bayesian Inference*, 2nd edn. London: Edward Arnold.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities.* Chichester: John Wiley & Sons.

Parham, F. and Portier C. J. (2005). Benchmark dose approach. In Edler, L. and Kitsos, C. (eds.), *Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment*, 239–254. Chichester: John Wiley & Sons.

Piegorsch, W. W. (1994). Environmental biometry: Assessing impacts of environmental stimuli via animal and microbial laboratory studies. In G. P. Patil and C. R. Rao (eds), *Handbook of Statistics: Environmental Statistics*, **12**, 535–559, Amsterdam: North Holland/Elsevier.

Piegorsch, W. W. and Bailer, A. J. (2005). *Analyzing Environmental Data*. Chichester: John Wiley & Sons.

Piegorsch, W. W., Xiong, H., Bhattacharya, R. N. and Lin, L. (2012). Nonparametric estimation of benchmark doses in quantitative risk analysis. *Environmetrics* **23**, 717–728.

Piegorsch, W. W., An, L., Wickens, A. A., West, R. W. Peña, E. A. and Wu, W. (2013). Information-theoretic model-averaged benchmark dose analysis in environmental risk assessment. *Environmetrics*, **24**, 143–157.

Portier, C. J. (1994). Biostatistical issues in the design and analysis of animal carcinogenicity experiments. *Environmental Health Perspectives* **102**, Suppl. 1, 5–8.

R Development Core Team. (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd Edn. New York: Springer-Verlag.

Robert, C. P. and Casella, G. (2011). A history of Markov chain Monte Carlo: subjective recollections from incomplete data. *Statistical Science* **26**, 102–115.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of computational and graphical statistics*, **18**, 349–367.

Sand, S., Victorin, K. and Falk Filipsson, A. (2008). The current state of knowledge on the use of the benchmark dose concept in risk assessment. *Journal of Applied Toxicology* **28**, 405–421.

Shao, K. (2012). A comparison of three methods for integrating historical information for Bayesian model averaged benchmark dose estimation. *Environmental Toxicology and Pharmacology* **34**, 288–296.

Shao, K. and Gift, J. S. (2014). Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. *Risk Analysis* **34**, 101–120.

Shao, K. and Small, M. J. (2011). Potential uncertainty reduction in model-averaged benchmark dose estimates informed by an additional dose study. *Risk Analysis* **31**, 1561–1575.

Shao, K. and Small, M. J. (2012). Statistical evaluation of toxicological experimental design for Bayesian model averaged benchmark dose estimation with dichotomous data. *Human and Ecological Risk Assessment* **18**, 1096–1119.

Simmons S. J., Chen C., Li X., Wang Y., Piegorsch W. W., Fang Q., Hu B. and Dunn G. E. (2013). Bayesian model averaging for benchmark dose estimation. Submitted.

Stern, A. H. (2002). Risk assessment, quantitative, In A. H. El-Shaarawi and W. W. Piegorsch (eds), *Encyclopedia of Environmetrics*, , **2**, 1025–1035, Chichester: John Wiley & Sons.

Stern, A. H. (2008). Environmental health risk assessment. In Melnick, E. L. and Everitt, B. S. (eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment* **2**, 580–589. Chichester: John Wiley & Sons.

Sun, D. and Tsutakawa, R. K. (1997). Bayesian design for dose-response curves with penalized risk. *Biometrics* **53**, 1262–1283.

U.S. EPA (2012). *Benchmark Dose Technical Guidance Document.* Technical Report number EPA/100/R-12/001. Washington, DC: U.S. Environmental Protection Agency.

U.S. General Accounting Office (2001). *Chemical Risk Assessment. Selected Federal Agencies' Procedures, Assumptions, and Policies.* Report to Congressional Requesters number GAO-01-810. Washington, DC: U.S. General Accounting Office.

U.S. National Toxicology Program (2009). *Toxicology and Carcinogenesis Studies of Cumene (CAS NO. 98-82-8) in F344/N Rats and B6C3F$_1$ Mice.* Technical Report number 542. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service.

Varadhan, R. and Gilbert, P. (2009). BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software* **32**, 1–26.

Vihola, M. (2012) Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing* **22**, 997–1008.

Wang, I. J., Chen, C. C., Chan, C. C., Chen, P. C., Leonardi, G. and Wu, K. Y. (2011). A hierarchical Bayesian approach for risk asessment of melamine in infant formula based on cases of related nephrolithiasis in children. *Food Additives and Contaminants* **28**, 384–395.

Wakefied, J. (2008). Geographic disease risk, In E. L. Melnick and B. S. Everitt (eds), *Encyclopedia of Quantitative Risk Analysis and Assessment*, Chichester: John Wiley & Sons, **2**, 744–748.

West, R. W., Nitcheva, D. K. and Piegorsch, W. W. (2009). Bootstrap methods for simultaneous benchmark analysis with quantal response data. *Environmental and Ecological Statistics* **16**, 63–73.

West, R. W., Pigorsch, W. W., Peña, E. A., An, L., Wu, W., Wickens, A. A., Xiong, H., Chen, W. (2012). The impact of model uncertainty on benchmark dose estimation. *Environmetrics* **23**, 706–716.

Wheeler, M. W. and Bailer, A. J. (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation. *Environmental and Ecological Statistics* **16**, 37–51.

Wheeler, M. W. and Bailer, A. J. (2009). Benchmark dose estimation incorporating multiple data sources. *Risk Analysis* **29**, 249–256.

Wheeler, M. W. and Bailer, A. J. (2012). Monotonic Bayesian semiparametric benchmark dose analysis. *Risk Analysis* **32**, 1207–1218.

Wickens, A. A. (2011). Model Adequacy in Benchmark Risk Analysis for Cancer Risk Assessment. Unpublished M. S. Thesis, University of Arizona, Tucson AZ.

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous response models, *Journal of Econometrics*, **25**, 365–393.