Dose-Response Modeling Under Simple Order Restrictions Using Bayesian Variable Selection Methods

Peer-reviewed author version

OTAVA, Martin; SHKEDY, Ziv; Lin, Dan; GOEHLMANN, Hinrich W.H.; BIJNENS, Luc; TALLOEN, Willem & KASIM, Adetayo (2014) Dose-Response Modeling Under Simple Order Restrictions Using Bayesian Variable Selection Methods. In: Statistics in Biopharmaceutical Research, 6 (3), p. 252-262.

DOI: 10.1080/19466315.2013.855472
Handle: http://hdl.handle.net/1942/17144

Dose-Response Modeling Under Simple Order Restrictions Using Bayesian Variable Selection Methods

Martin Otava, Ziv Shkedy

Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat),

Center for Statistics, Universiteit Hasselt, Belgium

Dan Lin

VMRD, Pfizer Animal Health, Zaventem, Belgium

Hinrich W.H. Göhlmann, Luc Bijnens and Willem Talloen

Janssen, Pharmaceutical companies of Johnson & Johnson, Beerse, Belgium

and

Adetayo Kasim

Wolfson Research Institute for Health and Wellbeing, Durham University, UK

Abstract

Bayesian modeling of dose-response data offers the possibility to establish the relationship between a clinical or a genomic response and increasing doses of a therapeutic compound and to determine the nature of the relationship wherever it exists. In this paper we focus on an order restricted one-way ANOVA model which can be used to test the null hypothesis of no dose effect against an ordered alternative. Within the framework of the dose-response modeling, a model uncertainty can be addressed using model averaging techniques. In this setting, the uncertainty is related to the number of all possible models that can be fitted to the data and should be taken into account for both inference and estimation. In this paper, we propose an order restricted Bayesian variable selection model that addresses the model uncertainty and can be used for both inference and estimation. The proposed method is applied to two case studies and is compared to the likelihood ratio test and the multiple contrast tests in both the analyses of the case studies and a simulation study.

Keywords: Bayesian modeling; Model uncertainty; Multiple contrast test; Order restricted models.

Acknowledgment: Martin Otava and Ziv Shkedy gratefully acknowledge the support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

1 INTRODUCTION

Dose-response experiments are an important part of a biomedical research to study relationships between increasing doses of a therapeutic compound and a variety of responses. Typically, the response represents a phenotypical effect of a compound such as inhibition, stimulation, toxicity, or expression level of a certain gene. The primary goal of such an experiment is to detect a dose-response relationship and to determine the nature of the relationship wherever it exists. In this paper, we focus on a continuous response and an experimental design with fixed number of doses. We further assume that the dose-response relationship, if exists, is monotone, i.e., the compound effect (increasing or decreasing) becomes stronger (or stays the same) with an increasing dose. Such property is very common in real applications, especially when inhibition or toxicity is measured. More general umbrella-shaped profiles (Bretz and Hothorn 2003) can occur within a context of an over-dosing and therefore a decreasing (increasing) effect is expected after reaching some threshold dose. This setting will not be considered further in the paper.

There are two main approaches for the analysis of dose-response experiments. The first approach uses parametric nonlinear models in order to estimate the dose-response relationship (Pinheiro et al. 2006; Whitney and Ryan 2009). The second approach assumes an underlying one-way ANOVA model with order restricted parameters (Robertson et al. 1988; Bretz and Hothorn 2003; Peddada et al. 2005; Lin et al. 2012) and can be used in order to test the null hypothesis of no dose effect against an ordered alternative.

We consider the second setting. The response is measured in $K-1$ dose levels and a control dose (placebo). Let μ_0 be the mean response under the control dose and $\mu_1, \mu_2, \ldots, \mu_{K-1}$ represent the mean responses under increasing doses of a therapeutic compound with $K-1$ dose levels. The primary interest is to test the null hypothesis of no dose effect given by

$$H_0 : \mu_0 = \mu_1 = \mu_2 = \ldots = \mu_{K-1}, \tag{1}$$

against an ordered alternative

$$H^{up} : \mu_0 \leq \mu_1 \leq \mu_2 \leq \ldots \leq \mu_{K-1}, \quad \text{or} \quad H^{dn} : \mu_0 \geq \mu_1 \geq \mu_2 \geq \ldots \geq \mu_{K-1}, \tag{2}$$

with at least one strict inequality. Here, H^{up} and H^{dn} correspond to an upward and downward directions of the order constraints, respectively. Note that the nature of the dose-response relationship, i.e., the dose-response curve shape, depends on a decomposition of the alternative hypothesis into a finite set of simpler hypotheses under monotone constraints. The decomposition results in $2^K - 1$ basic models under each of the monotone directions. For example, for a dose-response experiment with one control dose and three increasing dose levels (i.e., $K = 4$), the alternative hypothesis can be decomposed into seven basic models (Table 1). Denote the whole set of models as $\mathcal{G}_R$. The problem of an estimation of dose-response profile is therefore equivalent to the selection of monotone models that best describe the data given $\mathcal{G}_R$. When one particular model is selected and inference is done under the selected model, the uncertainty due to the model selection is ignored (Claeskens and Hjort 2008). Such an approach can lead to a bias in estimation of dose-specific means, especially when two models are almost equally supported by data. Inference within the dose-response framework that accounts for the model uncertainty is discussed by Pinheiro et al. (2006), Bornkamp et al. (2009) and Whitney and Ryan (2009) who use parametric nonlinear models to characterize the dose-response relationship and propose to use model averaging techniques to account for model uncertainty. Within this approach several parametric nonlinear models are fitted to the data and information from all models is combined, using information criteria, for both estimation and inference.

The likelihood ratio test (LRT, Barlow et al. 1972 and Robertson et al. 1988) and the multiple contrast tests (MCT, Mukerjee et al. 1987) are commonly use to test the null hypothesis of no dose effect. However, the inference of these tests ignores the uncertainty due to possible models that can be fitted to the data. In fact, the inference for the LRT is based on one specific model from all models under the alternative (the one that is esti-

Table 1: The set of eight possible monotonic dose-response models for an experiment with four dose levels (including placebo). Denote μ_i the mean response of dose level. The model g_0 represents the null model of no dose effect.

Model	Up: Mean Structure	Down: Mean Structure
g_0	$\mu_0 = \mu_1 = \mu_2 = \mu_3$	$\mu_0 = \mu_1 = \mu_2 = \mu_3$
g_1	$\mu_0 < \mu_1 = \mu_2 = \mu_3$	$\mu_0 > \mu_1 = \mu_2 = \mu_3$
g_2	$\mu_0 = \mu_1 < \mu_2 = \mu_3$	$\mu_0 = \mu_1 > \mu_2 = \mu_3$
g_3	$\mu_0 < \mu_1 < \mu_2 = \mu_3$	$\mu_0 > \mu_1 > \mu_2 = \mu_3$
g_4	$\mu_0 = \mu_1 = \mu_2 < \mu_3$	$\mu_0 = \mu_1 = \mu_2 > \mu_3$
g_5	$\mu_0 < \mu_1 = \mu_2 < \mu_3$	$\mu_0 > \mu_1 = \mu_2 > \mu_3$
g_6	$\mu_0 = \mu_1 < \mu_2 < \mu_3$	$\mu_0 = \mu_1 > \mu_2 > \mu_3$
g_7	$\mu_0 < \mu_1 < \mu_2 < \mu_3$	$\mu_0 > \mu_1 > \mu_2 > \mu_3$

mated by the isotonic regression) while inference for MCT takes into account that different models are possible under the alternative but the inference is done based on one specific model as well. In this paper, we focus on Bayesian variable selection (BVS) models for order restricted one-way ANOVA models for dose-response data that offers a framework to simultaneously establish a dose-response relationship (inference) and to determine the nature of the relationship (estimation). For both inference and estimation, model uncertainty is taken into account. The proposed BVS approach is closely related to the Gibbs variable selection method proposed by Whitney and Ryan (2009) for the case of parametric nonlinear dose-response modeling. However, in contrast with the Gibbs variable selection approach, the BVS approach estimates the posterior probability for each one of the models in $\mathcal{G}_R$. The posterior mean response at each dose level is a weighed average of the posterior means of all models in which the weights are the model posterior probability. In addition, the posterior probability of the null model is of the primary interest, since it also represents a probability for false positives, i.e., wrongly rejecting the null hypothesis, and therefore can be used for the inference (Newton et al. 2007).

The paper is organized as follows. In Section 2, we describe the case studies used for the analysis presented in the manuscript. The current frequentist procedures, formulation of the hierarchical Bayesian model for the dose-response data and the Bayesian variable selection approach are discussed in Section 3. In Section 4, we present the results from the application of the methodology to the case studies. The simulation study for the comparison between BVS and the frequentist methods is introduced in Section 5 and the paper is concluded with a discussion in Section 6.

2 CASE STUDIES

Two real life studies are used to illustrate the methodology discussed in this paper. Both of the case studies have the same data structure: response is measured under increasing doses of the respective compounds with the first dose being a control (placebo). The dose-response data for the case studies are shown in Figure 1. The third case study is presented in he supplementary appendix of the manuscript.

2.1 The AMES Study

The AMES dataset (Bretz and Hothorn 2003) contains the data about a mutagenicity level of a compound. The mutagenicity is reflected by an increasing relationship between dose level and frequency of visible colonies among plated salmonella bacteria. Dose level is used as a covariate and a frequency of colonies as a response. Although we suspect very high doses to lower number of microbes due to toxicity, in the following analysis we assume only the nondecreasing profile. More detailed information about the data can be found in Ames et al. (1975). Five observations are available for a placebo and three for each of four active doses.

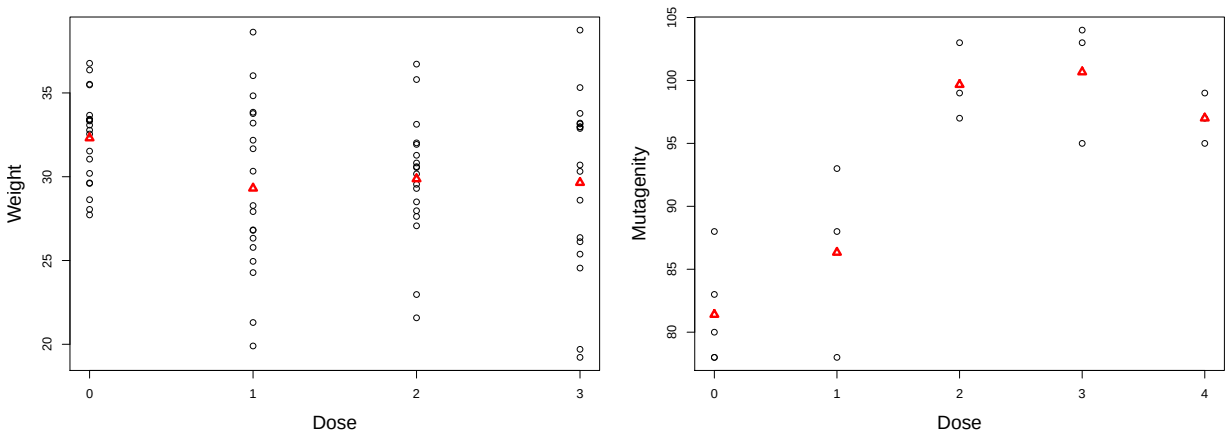

Figure 1: *Case studies analyzed in the manuscript. Triangles represent dose-specific means. Left panel: the LITTER dataset. Right panel: the AMES dataset.*

2.2 The LITTER Study

The LITTER dataset is available in R (R Core Team 2013) as part of the `R` package `multcomp` (Hothorn et al. 2008). It contains data about pregnant mice that were divided into four groups and the compound in four different doses was administered during pregnancy. For a placebo, 20 mice were used, for active doses 19, 18 and 17 mice, respectively. The litters were evaluated for birth weights. We focus on relationship between the birth weight and the dose. For the LITTER dataset, the null hypothesis of no dose effect is tested against the nonincreasing alternative in order to detect toxicity effects due to the used drug.

3 BAYESIAN VARIABLE SELECTION MODELS FOR DOSE-RESPONSE MODELING

3.1 Testing The Null Hypothesis Against a Simple Ordered Alternative

The basic setting we considered in this paper consists of a response variable measured in a sequence of dose levels. Let Y_{ij} represents the response for jth observation at the dose level i and μ_i denotes the mean response at the dose level i. In order to model the relationship between the response and the increasing doses of a therapeutic compound we formulate the following linear model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \tau^{-1}), \quad i = 0, \ldots, K-1, \quad j = 1, 2, \ldots, n_i \tag{3}$$

For a given direction, the likelihood ratio test (LRT) computes the maximum likelihood estimates for the mean response under the two hypotheses formulated in (2). The maximum likelihood estimator computed under the null hypothesis H_0 equals the sample mean $\hat{\mu} = \left(\sum_{i=0}^{K-1} \sum_{j=1}^{n_i} Y_{ij}\right) / \sum_{i=0}^{K-1} n_i$. The maximum likelihood estimator under the order restricted alternative H^{up} is the vector of isotonic means (Robertson et al. 1988). The likelihood ratio test statistic, proposed by Barlow et al. (1972), can be expressed as

$$T_{LRT} = \frac{RSS_0 - RSS_1}{RSS_0} = 1 - \frac{RSS_1}{RSS_0}, \tag{4}$$

where RSS_0 represents the residual sum of squares under the null hypothesis and RSS_1 the residual sum of squares under the alternative H^{up} (or H^{dn}). The null hypothesis is rejected for a large value of T_{LRT} (details about procedure are given in a supplementary appendix of the manuscript).

A second approach to test the null hypothesis is the so-called multiple contrast test (MCT). The motivation for developing multiple contrast tests by Mukerjee et al. (1987) was to achieve

tests with similar power to the LRT, but easier to use and interpret (Lin et al. 2012). The key idea is to perform as small number of comparisons as possible while covering sufficiently the space of all possible alternative hypotheses. The test is based on the simultaneous testing of the multiple comparisons using the contrast test defined as

$$T^{SC} = \frac{\sum_{i=0}^{K-1} c_i \hat{\mu}_i}{s \cdot \sqrt{\sum_{i=0}^{K-1} \frac{c_i^2}{n_i}}}, \tag{5}$$

where $\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$, $s = \sqrt{\frac{1}{\nu} \sum_{i=0}^{K-1} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2}$ and $\nu = \sum_{i=0}^{K-1} (n_i - K)$.

The contrast vector $\boldsymbol{c} = (c_0, \ldots, c_{K-1})$ fulfills the condition $\sum_{i=0}^{K-1} c_i = 0$. Bretz (2006) shows that, under normality assumption, the test statistic T^{SC} follows an univariate central t-distribution with ν degrees of freedom under H_0. The MCT test statistic is the maximum over these V single contrast tests:

$$T^{MC} = \max_{v=1,\ldots,V} \{T_1^{SC}, T_2^{SC} \ldots T_V^{SC}\}. \tag{6}$$

Covering the space of the alternative hypotheses translates into a choice of a combination of vectors $\boldsymbol{c}_v$, $v = 1, \ldots, V$ (Lin et al. 2012). The choice of the set of the vectors $\boldsymbol{c}_v$ determines properties of the test and distinguish between the different MCTs (Hothorn 2006). For further comparison, Williams' and Marcus' MCTs (Bretz 1999) based on the tests proposed by Williams (1971) and Marcus (1976) are used. Williams' MCT is based on the comparison of the last dose mean $\hat{\mu}_{K-1}^{\star}$, estimated by the isotonic regression, with $\hat{\mu}_0$, the estimate of the mean of the first dose, under the different possible profiles. Marcus' MCT is a modification of Williams' MCT in which $\hat{\mu}_0$ is replaced with the isotonic estimate $\hat{\mu}_0^{\star}$. An elaborate discussion about both the LRT and Williams' and Marcus' MCTs is given in the supplementary appendix of the manuscript.

3.2 Bayesian Estimation Under Strict Inequality Constraints

Our aim is to estimate the parameters under a strict inequality constraints $\mu_0 < \mu_1 < \mu_2 < \cdots < \mu_{K-1}$. The constraints can be achieved by constraining the parameter space of $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_{K-1})$, whereby the order restrictions are imposed on the prior distributions. For a monotone upward profile we assume that for a profile function $\psi(i)$ it holds that $\psi(i) = \mu_{\lfloor i \rfloor}$ and that $\psi(i)$ is a right-continuous, nondecreasing function defined on interval $[0, K-1]$. We do not assume any deterministic relationship between μ_i and the dose levels, instead we specify a probabilistic model for μ_i at each distinct dose level.

To estimate $\boldsymbol{\mu}$ under the order restrictions, $\mu_0 < \mu_1 < \ldots < \mu_{K-1}$, the K dimensional parameter vector is constrained to lie in a subset $S^K \in \mathbb{R}^K$. The constrained set S^K is determined by the order among the components of $\boldsymbol{\mu}$. In this case, it is natural to incorporate the constraints into the specification of the prior distribution (Klugkist and Mulder 2008). Let $\boldsymbol{Y} = (Y_{11}, Y_{12}, \ldots, Y_{K-1,n_{K-1}})$ be the response value and $\boldsymbol{\eta}$ and $\boldsymbol{\tau}$ the hyper parameters for $\boldsymbol{\mu}$. Gelfand et al. (1992) showed that the posterior distribution of $\boldsymbol{\mu}$, given the constraints, is the unconstrained posterior distribution normalized such that

$$P(\boldsymbol{\mu}|\boldsymbol{Y}) \propto \frac{P(\boldsymbol{Y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau})}{\int_{S^k} P(\boldsymbol{Y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau})d\boldsymbol{\mu}}, \quad \boldsymbol{\mu} \in S^K. \tag{7}$$

Let $S_i^K(\mu_l, l \neq i)$ be a cross section of S^K defined by the constraints for μ_i at a specified set of μ_l, with $l = 0, 1, 2, \ldots, i-1, i+1, \ldots, K-1$. In our setting, $S_i^K(\mu_l, l \neq i)$ is part of the interval $[\mu_{i-1}, \mu_{i+1}]$. It follows from (7) that the posterior distribution for μ_i is given by

$$\begin{cases} P(\mu_i|\boldsymbol{Y},\boldsymbol{\eta},\boldsymbol{\tau},\boldsymbol{\mu}_{-i}) \propto P(\boldsymbol{Y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau}), & \mu_i \in S_i^K(\mu_l, l \neq i), \\ 0, & \mu_i \notin S_i^K(\mu_l, l \neq i). \end{cases} \tag{8}$$

where $\boldsymbol{\mu}_{-i} = (\mu_0, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_{K-1})$. Hence, when the likelihood and the prior distribution are combined, the posterior conditional distribution of $\mu_i|\boldsymbol{Y},\boldsymbol{\eta},\boldsymbol{\tau},\boldsymbol{\mu}_{-i}$ is the standard posterior distribution restricted to $S_i^K(\mu_l, l \neq i)$, i.e., restricted to the interval $[\mu_{i-1}, \mu_{i+1}]$

(Gelfand et al. 1992). As a result, during the MCMC simulations, the sampling from the full conditional distribution can be reduced to the interval restricted sampling from the standard posterior distribution. Following Klugkist and Mulder (2008), we formulate an order restricted ANOVA model for which the mean response at the ith dose level is given by

$$E(Y_{ij}) = \mu_i = \begin{cases} \mu_0, & i = 0, \\ \mu_0 + \sum_{\ell=1}^{i} \delta_\ell, & i = 1, \ldots, K-1 \end{cases} \tag{9}$$

with the constraints that $\delta_\ell \geq 0$ for an upward trend or $\delta_\ell \leq 0$ for a downward trend. In a matrix notation, the mean gene expression for an upward trend model (for $K = 4$) is given by

$$E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{cases} \mu_0, & \text{control}, \\ \mu_0 + \delta_1, & \text{first dose level}, \\ \mu_0 + \delta_1 + \delta_2, & \text{second dose level}, \\ \mu_0 + \delta_1 + \delta_2 + \delta_3, & \text{third dose level}. \end{cases} \tag{10}$$

In order to complete the specification of the hierarchical model, we assume the following prior distribution for the unknown model parameters,

$$\begin{aligned} &\mu_0 \sim TN(\eta_{\mu_0}, \tau_{\mu_0}^{-1}, 0, \infty), \\ &\delta_i \sim TN(\eta_{\delta_i}, \tau_{\delta_i}^{-1}, 0, A) \qquad k =, 1, \ldots, K-1. \end{aligned} \tag{11}$$

Here $TN(\mu, \sigma^2, a, b)$ is a truncated normal distribution and A is a positive constant. The truncated distribution improves properties of the MCMC chains and it is a priori constrained

to lie between zero and the difference in the range of the response vector. We assume a noninformative distribution for the hyper parameters in the model,

$$\begin{aligned}
&\tau \sim \Gamma(10^{-3}, 10^{-3}),\\
&\eta_{\mu_0} \sim N(0, 10^6),\\
&\tau_{\mu_0} \sim \Gamma(10^{-3}, 10^{-3}),\\
&\eta_{\delta_k} \sim N(0, 10^6), k = 1, \ldots, K-1\\
&\tau_{\delta_k} \sim \Gamma(10^{-3}, 10^{-3}), k = 1, \ldots, K-1.
\end{aligned} \tag{12}$$

3.3 Bayesian Variable Selection Models

The Bayesian inequality model defined above cannot be used in our framework due to the equality constraints on the means of the null model and some of the alternative models. As pointed out by Dunson and Neelon (2003), since the priors of the components of $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_{K-1})$ are the truncated normal distributions, the mean structure $\mu_i = \mu_0 + \sum_{\ell=1}^{i} \delta_\ell$ implies an order constraints mean structure with the strict inequalities $\mu_0 < \mu_1 <, \ldots, < \mu_{K-1}$. The equality constraints would, in practice, assign zero probabilities to all other competing models except the model with the strict inequality constraints (Klugkist and Hoijtink 2007). In what follows we propose a Bayesian variable selection model that can be seen as an extension of the informative hypothesis inference framework discussed by Klugkist and Hoijtink (2007) to the setting in which equality constraints can be incorporated in the mean structure. Then, all the different models under the alternative hypothesis are taken into account for both inference and estimation. The equality constraints can be incorporated in the model by setting some of the components in $\boldsymbol{\delta}$ to be equal to zero. Indeed, $\delta_i = 0$ implies $\mu_i = \mu_{i-1}$.

The differences in the mean structures of the different models, therefore, depends on which of the components in $\boldsymbol{\delta}$ are set to be equal to zero or equivalently which columns in the ordered design matrix $\boldsymbol{X}$ are excluded. Hence, the design matrix for each of the models is in

fact a subset of the design matrix $\boldsymbol{X}$. As a result, the problem of the model estimation in the presence of equality constraints is reduced to a problem of the variable selection depending on which of the columns of $\boldsymbol{X}$ are selected or deleted. This is related to the Bayesian variable selection approach (George and McCulloch 1993) which is used to determine an optimal model from a priori set of R known plausible models. As pointed out by O'Hara and Sillanpää (2009) the choice of an optimal model reduces to the choice of a subset of variables which are included in the model (i.e., model selection), or the choice of which parameters in the parameter vector are different from zero (i.e., inference). Let z_i, $i = 1, \ldots, K$ be an indicator variable such that

$$z_i = \begin{cases} 1, & \delta_i \quad \text{is included in the model,} \\ 0, & \delta_i \quad \text{is not included in the model,} \end{cases}$$

and let $\theta_i = \delta_i \times z_i$, $\boldsymbol{\beta} = (\mu_0, \theta_1, \theta_2, \theta_3)$ and $\boldsymbol{Y}$ be the response vector. Hence, we can reformulate the mean structure in (9) (O'Hara and Sillanpää 2009) in terms of θ_i and z_i as

$$E(Y_{ij}) = \mu_0 + \sum_{\ell=1}^{i} \theta_\ell = \mu_0 + \sum_{\ell=1}^{i} z_\ell \delta_\ell. \tag{13}$$

For the four dose level experiment ($K = 4$) discussed above, the triplet $\boldsymbol{z} = (z_1, z_2, z_3)$ defines uniquely each one of the 8 plausible models. For example, for $\tilde{\boldsymbol{z}}_1 = (0, 0, 0)$ holds that $E(Y_{ij}|\mathcal{G}_R, \boldsymbol{z} = \tilde{\boldsymbol{z}}_1) = (\mu_0, \mu_0, \mu_0, \mu_0)$ (which corresponds to the mean of the model g_0) and for $\tilde{\boldsymbol{z}}_2 = (1, 0, 0)$ we obtain $E(Y_{ij}|\mathcal{G}_R, \boldsymbol{z} = \tilde{\boldsymbol{z}}_2) = (\mu_0, \mu_0 + \delta_1, \mu_0 + \delta_1, \mu_0 + \delta_1)$ (which corresponds to the mean of the model g_2). Hence, in our setting the BVS model estimates the posterior probability of each model, $P(g_r|\text{data})$ and in particular the posterior probability of the null model $P(g_0|\text{data})$. For example, $P[\boldsymbol{z} = (0, 0, 0)|\text{data}] = P[E(Y_{ij}) = \mu_0|\text{data}]$.

Kuo and Mallick (1998) approach was used for the specification of the prior models for z_i and δ_i. It assumes that z_i and δ_i are independent, i.e., $P(\delta_i, z_i) = P(\delta_i) \times P(z_i)$, with a truncated normal prior distribution for δ_i (Equation 11) and

$$\begin{aligned} z_i &\sim \text{Bernoulli}(\pi_i), \\ \pi_i &\sim \text{U}(0, 1). \end{aligned} \tag{14}$$

The variable π_i represents inclusion probability of z_i and can be estimated by the proportion of the $z_i = 1$ within the MCMC run.

As pointed out by O'Hara and Sillanpää (2009), the posterior inclusion probability of δ_i in the model is the posterior mean of z_i. Further, for a given value of K, using the indicator variable z_i we specify a transformation function that uniquely defines each one of the plausible models (Ntzoufras 2002), $G = 1 + \sum_{i=1}^{K-1} z_i 2^{i-1}$. Thus, the posterior probability of $G = r+1$ defines uniquely the posterior probability of a specific model g_r (when g_r defined as in Table 1). In particular (for K=4), the posterior probability of the null model is given by

$$P(G = 1|\text{data}) = P\left[E(Y_{ij}) = \mu_0|\text{data}\right] = P\left[\bar{\boldsymbol{z}} = c(0,0,0)|\text{data}\right] = P(g_0|\text{data}). \tag{15}$$

For $K = 4$ there are eight possible monotone models (for a given direction): seven monotone models (given in Table 1) and the null model. It follows that G is given by

$$G = \begin{cases} 1, & \text{for } \boldsymbol{z} = (z_1 = 0, z_2 = 0, z_3 = 0), \quad \text{model } g_0, \\ 2, & \text{for } \boldsymbol{z} = (z_1 = 1, z_2 = 0, z_3 = 0), \quad \text{model } g_1, \\ 3, & \text{for } \boldsymbol{z} = (z_1 = 0, z_2 = 1, z_3 = 0), \quad \text{model } g_2, \\ 4, & \text{for } \boldsymbol{z} = (z_1 = 1, z_2 = 1, z_3 = 0), \quad \text{model } g_3, \\ 5, & \text{for } \boldsymbol{z} = (z_1 = 0, z_2 = 0, z_3 = 1), \quad \text{model } g_4, \\ 6, & \text{for } \boldsymbol{z} = (z_1 = 1, z_2 = 0, z_3 = 1), \quad \text{model } g_5, \\ 7, & \text{for } \boldsymbol{z} = (z_1 = 0, z_2 = 1, z_3 = 1), \quad \text{model } g_6, \\ 8, & \text{for } \boldsymbol{z} = (z_1 = 1, z_2 = 1, z_3 = 1), \quad \text{model } g_7. \end{cases} \tag{16}$$

The BVS model discussed in this section is not only an extension of the informative hypothesis inference framework of Klugkist and Hoijtink (2007), but provides in addition a model selection procedure that can be based on the maximum posterior model probability. Furthermore, the posterior mean at each dose level can be seen as a Bayesian model average of all the models that were fitted during the MCMC simulation. The weights for the Bayesian model average are the posterior model probabilities $P(G = r|data, g_0, \ldots, g_R)$. Thus, the BVS model provides a framework for inference, model selection and estimation that takes

into account model uncertainty.

4 APPLICATION TO THE CASE STUDIES

The dataset from each study was analyzed using the LRT, the MCT with Williams' and Marcus' contrast and the BVS model. The BVS models were fitted in `Winbugs 1.4` using MCMC simulation with 20,000 iterations from which the first 5,000 were discarded as burn-in period.

4.1 The AMES Data

The results obtained for all the methods are presented in Table 2. All the frequentist methods show no evidence in favour of the null hypothesis. The posterior probability of the null model obtained for the BVS model ($6.7 \cdot 10^{-5}$) indicates no evidence in favour of the null model, but substantive evidence in support of an alternative model with monotone relationship between the frequency of mutation and the increasing doses of the compound (0.408).

Figure 2a reveals a close agreement between the posterior means obtained for the BVS model and maximum likelihood parameter estimates obtained by the isotonic regression for the AMES study. Note that the posterior means obtained from the BVS model do not correspond to the one specific model but it is the Bayesian weighted model averaging of all competing models (for $K = 5$ there are 16 possible models, including the null model). Interestingly, similar to the isotonic regression which pools together the means of the last three dose levels, the inclusion probabilities (Figure 2b) obtained from the BVS model show little evidence in support of different dose effects for dose 3 and dose 4 (with the estimated posterior probabilities of 0.11 and 0.09, respectively).

Table 2: P-values for the frequentist methods and the posterior model probabilities for the BVS model. "BVS null" shows the posterior probability of the null model and "BVS max" shows the maximal posterior probability among the posterior probabilities of all the alternative monotone models.

	LRT	MCT(W)	MCT(M)	BVS null	BVS max
AMES	$6 \cdot 10^{-5}$	$1.4 \cdot 10^{-5}$	$3.6 \cdot 10^{-5}$	$6.7 \cdot 10^{-5}$	0.408
LITTER	0.029	0.019	0.029	0.220	0.623

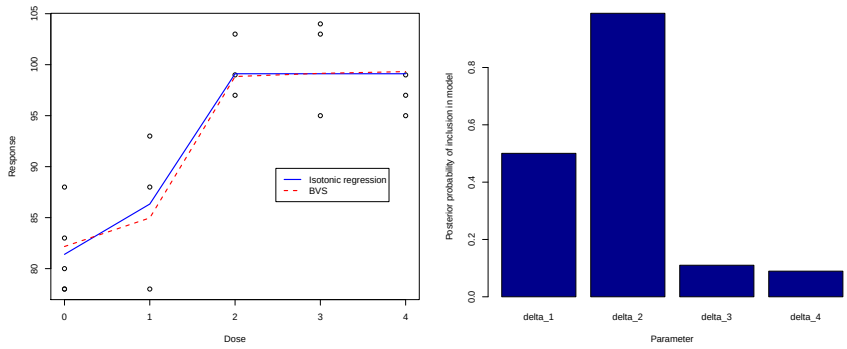

Figure 2: *The AMES mutagenity data. Left panel: observed data, isotonic regression (solid line) and posterior mean of the BVS model (dashed line). Dotted line; the posterior mean obtained by MCMC when only model with maximum posterior probability for BVS was taken into account. Right panel: posterior mean of the inclusion probability of δ_i into the model.*

4.2 The LITTER Data

The p-values and the posterior model probabilities for the LITTER data are shown in 2. The LRT and MCTs reject the null hypothesis. The posterior probability of the null hypothesis obtained from BVS is 0.22, which implies that there is more support in favor of the alternative hypothesis given the data. Specifically, the BVS shows more substantive evidence in support of the alternative model g_1 (defined in Table 1) whose posterior model probability is 0.623 (see Figure 3b). This model has a common dose effects for dose 1 to dose 3. This illustrates

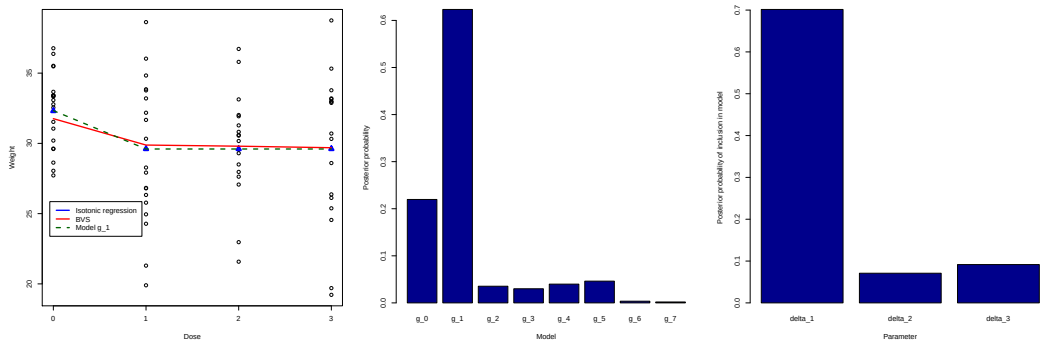

Figure 3: *The LITTER data. Left panel: observed data, isotonic regression (solid line) and posterior mean of the BVS model (dashed line). Dotted line; the posterior mean obtained by MCMC when only model g_1, i.e., model with maximum posterior probability for BVS, was taken into account. Dotted line coincides with solid line almost perfectly. Right panel: Posterior probability of null model g_0 and alternative models g_i, $i = 1, \ldots, 7$. Notation corresponds to the model numbers presented in Table 1.*

an important aspect of the BVS model which simultaneously performs the inference and provides the evidence for all the possible models given the data. Furthermore, the inclusion probabilities, shown in Figure 3c, indicate that the δ_2 and δ_3 should not be included in the model which corresponds to the results obtained from the isotonic regression.

Due to the fact that the posterior probability of model g_1 is relatively high compared to the other models, the posterior means of the BVS model are similar to those of the isotonic means and the posterior means from g_1 with the common mean for dose 1 to dose 3 and the different mean for control (Figure 3a). Note that model g_1 is different from the BVS model since its design matrix is fixed while the BVS fits all the possible models simultaneously and produce the model averaging of the posteriors means for doses across all the competing models, weighted by their respectively posterior probabilities given the data.

5 SIMULATION STUDY

A simulation study was conducted in order to investigate the performance of the BVS model in terms of controlling the Type I error and the power. The simulation settings correspond to the setting used by Marcus (1976) and experiment with four and five dose levels was investigated. Configurations for the mean response at each dose levels, $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2, \mu_3)$ and $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_4)$, for $K = 4$ and $K = 5$ are given in Table 12 and Table 13 in the supplementary appendix of the manuscript. For each simulation setting the data were generated according to the model $Y_{ij} = \lambda\mu_i + \varepsilon_{ij}$, $\varepsilon_{ij} \sim \mathrm{N}(0, 1)$. A sequence of $\lambda = 0, 1, 1.5, 2, 2.5, 3$ were used to investigate the magnitude of the differences between the mean response across the doses. Note that using $\lambda = 0$ implies the null model with no dose effect. Number of observations per dose was equal to $n = 3, 4, 5$.

The BVS model, one-sided LRT and one-sided MCTs were performed. Table 3 shows the empirical Type I error obtained from each method. All the methods control the Type I error, while the MCTs are more conservative than the LRT. The BVS model seems even more conservative. To achieve similar proportion of false rejections as in case of the LRT and the MCTs, i.e. 0.05, we can use a threshold as high as 0.35 for the BVS rejection (see Table 3 and Figure 5 in the supplementary appendix).

Table 4 shows the power of the methods for $K = 4$. As expected the the LRT seems to be the most powerful test with both MCTs slightly worse and BVS with threshold 0.05 about 0.10 behind the MCTs. With an increasing λ, the difference between the methods diminishes (this pattern is visualized in Figure 4 for $n = 4$). Such result is expected, because with higher λ, the power tends to one for all the methods. Similar result was obtained for the case of five dose levels. The improving performance of the BVS with an increasing threshold is natural, too. When using higher threshold (that still controls Type I error) we achieve results in terms of power comparable with frequentist methods (Figure 4). The results for the case of $n = 5$ observations per dose level, shown in Table 3 in the supplementary appendix of

Table 3: Type I error of the frequentist methods and the BVS model for $K = 4$ and $K = 5$.

	LRT	MCT(W)	MCT(M)	BVS 0.05	BVS 0.10	BVS 0.15	BVS 0.35
$K = 4$							
$n = 3$	0.041	0.037	0.036	0.002	0.003	0.003	0.034
$n = 4$	0.044	0.044	0.048	0.002	0.002	0.005	0.027
$n = 5$	0.053	0.057	0.051	0.001	0.001	0.001	0.017
$K = 5$							
$n = 3$	0.047	0.048	0.048	0.000	0.002	0.005	0.046
$n = 4$	0.056	0.051	0.052	0.000	0.002	0.003	0.030
$n = 5$	0.048	0.056	0.051	0.000	0.002	0.004	0.022

the manuscript, indicate that the BVS model achieves the similar power compared to the frequentist methods. The very same pattern can be seen for $K = 5$, tables with all the results are included in the supplementary appendix of the manuscript.

Figure 5 demonstrates visually the change in the power when the number of dose levels increase from $K = 4$ to $K = 5$ which corresponds to a change from 1/8 to 1/16 for the model prior probabilities, respectively. Note that the first seven models corresponds to $K = 4$ (circles) while the last 15 models corresponds to $K = 5$ (filled circles). We can see that a change in the power across the models and dose levels for the BVS model behaves in very similar way as the change in the power obtained for the LRT. Hence, one can conclude that the change in power is due to the additional information provided by the data (by adding more dose levels) and not due to the change in the prior probabilities. Similar patterns were observed for different values of λ and n (e.g. Figure 7, 8 in the supplementary appendix).

Table 4: Results for $K = 4$ and $n = 3$. The columns RT and MCTs show estimation of the power of the particular tests. The columns BVS shows proportion of (mean of) posterior probabilities of the null model given the data that are smaller then $\alpha = 0.05, 0.10, 0.15, 0.35$.

λ	Profile	LRT	MCT (W)	MCT (M)	BVS 0.05	BVS 0.10	BVS 0.15	BVS 0.35
1	g_1	0.36	0.42	0.34	0.22	0.37	0.49	0.81
	g_2	0.38	0.31	0.36	0.22	0.37	0.48	0.80
	g_3	0.40	0.39	0.35	0.22	0.38	0.50	0.83
	g_4	0.36	0.26	0.35	0.22	0.38	0.50	0.83
	g_5	0.44	0.42	0.39	0.26	0.41	0.54	0.86
	g_6	0.41	0.33	0.38	0.22	0.36	0.49	0.82
	g_7	0.46	0.42	0.41	0.24	0.40	0.52	0.85
2	g_1	0.85	0.90	0.85	0.74	0.88	0.93	0.99
	g_2	0.86	0.73	0.84	0.74	0.88	0.94	0.99
	g_3	0.89	0.88	0.86	0.81	0.90	0.95	0.99
	g_4	0.85	0.72	0.82	0.74	0.85	0.92	0.99
	g_5	0.90	0.91	0.87	0.80	0.92	0.96	1.00
	g_6	0.90	0.81	0.87	0.80	0.91	0.96	1.00
	g_7	0.90	0.88	0.87	0.82	0.93	0.97	1.00
3	g_1	0.99	1.00	0.99	0.98	0.99	1.00	1.00
	g_2	0.99	0.98	0.99	0.99	1.00	1.00	1.00
	g_3	0.99	0.99	0.99	0.98	0.99	1.00	1.00
	g_4	0.99	0.97	0.99	0.97	0.99	1.00	1.00
	g_5	1.00	1.00	0.99	0.98	1.00	1.00	1.00
	g_6	1.00	0.98	0.99	0.99	1.00	1.00	1.00
	g_7	1.00	0.99	0.99	0.99	1.00	1.00	1.00

6 DISCUSSION

In many applications, an analysis of the dose-response data requires to test the null hypothesis of no dose effect against an ordered alternative or to estimate the dose-response curve. In this paper we focus on a Bayesian approach for an order constrained one-way ANOVA models. The inequality constraints were incorporated as priors in the Bayesian formulation of the model. We have shown that the approach of (Gelfand et al. 1992) assigns zero probabilities to the models with equality constraints, which cannot be used in our setting. In order to overcome the problem of the zero probabilities for the quality constraints, we introduced

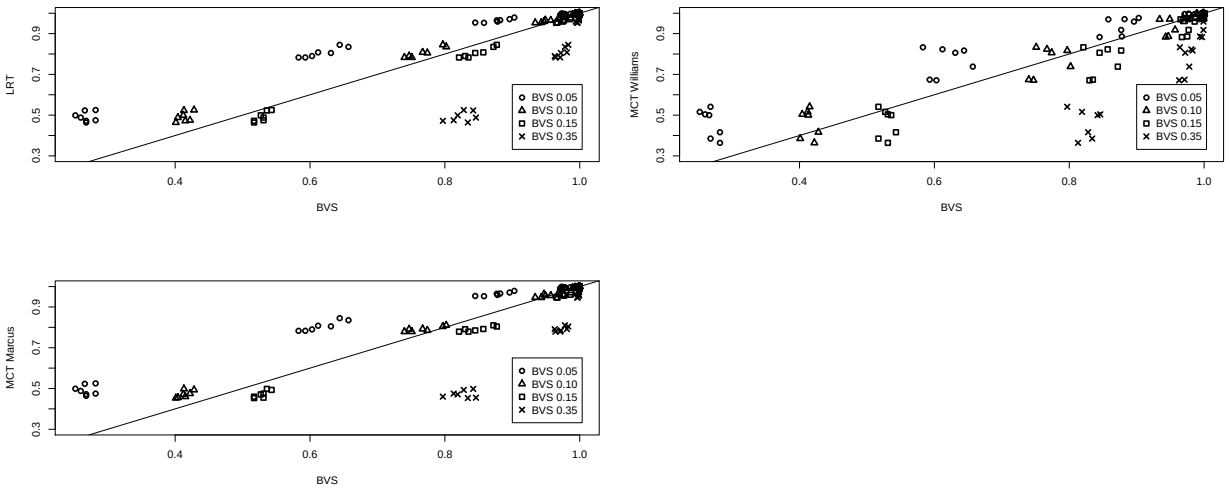

Figure 4: *Comparison of the power between the BVS (with varying threshold) and the frequentist tests. Circles represent the results for threshold* $\alpha = 0.05$*, triangles* $\alpha = 0.10$ *and rectangles* $\alpha = 0.15$*. The plot is based on simulation under* $n = 4$*. Top left: LRT vs. BVS. Top right: MCT Williams vs. BVS, Bottom left: MCT Marcus vs. BVS.*

the BVS model formulation for the dose-response modeling.

The BVS model presented in the manuscript assumes an independent prior model for the joint distribution of z_i and δ_i, i.e., $P(\delta_i, z_i) = P(\delta_i) \times P(z_i)$ and non-informative priors for both $P(\delta_i)$ and $P(z_i)$. An alternative approach is to formulate a model for $P(z_i, \delta_i)$ by taking into account the conditional distribution $P(\delta_i|z_i)$. Dellaportas et al. (2002) proposed the Gibbs Variable Selection (GVS) method which assumes a mixture model for the conditional distribution $P(\delta_i|z_i)$, i.e., $P(\delta_i|z_i) = z_i N(\eta_i, S) + (1 - z_i)N(0, \tau^2)$. The Stochastic Search Variable Selection (SSVS) by (George and McCulloch 1993) assumes the following mixture model for $P(\delta_i|z_i) = z_i N(0, \tau^2) + (1 - z_i)N(0, g\tau^2)$. In both GVS and SSVS is necessary to specify priors for the tuning parameters (S and τ^2 for GVS, g and tau^2 for SSVS). In both cases a prior knowledge about the increment is needed for specification. The influence of the choice of the prior model for $P(\delta_i|z_i)$ will be investigated further and will be reported in a separate manuscript.

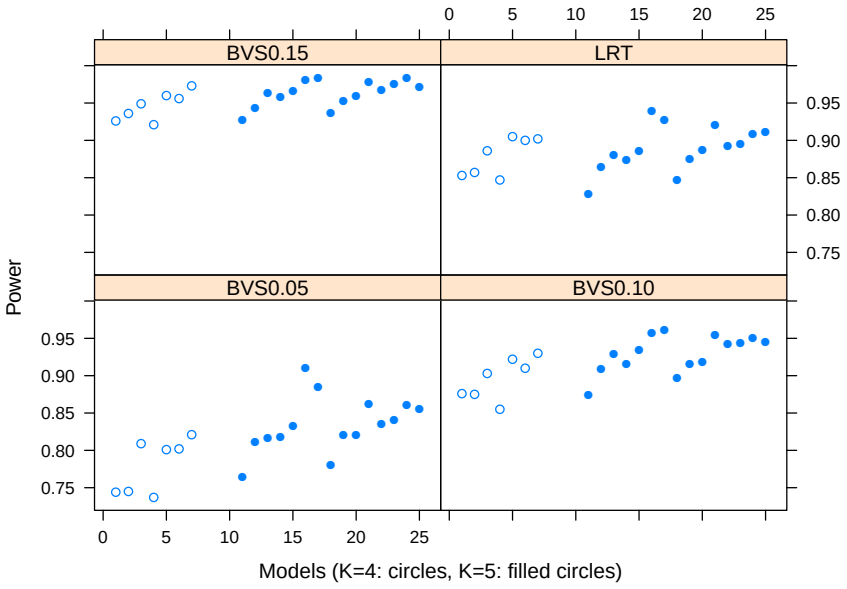

Figure 5: *Comparison of the power between $K = 4$ and $K = 5$ for BVS (with varying threshold) and LRT test (top right panel). The plot is based on a simulation under $\lambda = 2$ and $n = 3$. The models are ordered arbitrarily, seven models for $K = 4$ on the left (circles) and 15 models for $K = 5$ on the right (filled circles).*

We have shown that using the BVS model allows us to calculate the posterior probability for each one of the candidate models and in particular the posterior probability of the null model. Therefore, the BVS model proposed in this paper can be used for both the inference and the estimation of the dose-response curve. Further, the posterior mean obtained from the BVS model is a model average of all the candidate order restricted one-way ANOVA models for a given value of the dose levels.

The simulation study showed that the BVS can match the frequentist methods in terms of the power while controlling similar level of the Type I error. The comparison is valid,

because we avoid to compare p-values and the posterior probabilities themselves, but rather the results based on using any of these two quantities for answering a question of the null hypothesis testing. The power of the BVS method indeed depends on chosen threshold. The approach on how to choose or estimate the threshold for a real dataset is one of the topics planned for a future research.

Framework of the microarray data enables us to overcome a necessity of the threshold specification by using the FDR control mechanism. Details are given in Section Multiplicity adjustment in the supplementary appendix of the manuscript.

Additionally, the BVS provides an evidence for the possible models under the monotone constraints. The probability of identification of the true monotone profile based on the posterior probabilities may bring further insight into the BVS model properties. Together with the comparison between the posterior probability of the most likely model and the posterior probability of other models, the BVS may also be used for the model selection among the alternative monotone models. We suggest this problem as a topic for a further research.

The BVS model presented in the paper was based on the use of non-informative priors for the selection variables $z_1, \ldots, z_K$. Strong prior scientific knowledge is typically rare in dose-response modeling situations, but when it is present (e.g., if historical data are available), it can be very easily incorporated. Simple change of the hyperprior for π_i or even straightly prior for z_i immediately translates into change in the prior probability of the null hypothesis and all the alternative profiles. Indeed, such a change can highly influence the posterior probability of the different models and so the estimated dose-specific means (since they are in fact weighted average of model-specific means with weights equaled to the posterior probability of the models). Hence, we suggest to use informative priors only in cases, when scientific knowledge is really strong and to specify them very carefully. Analysis of the impact of particular changes of priors on posteriors is very complex, because it depends on multiple criteria such as properties of the observed data set or the number of dose levels.

References

Ames, B. N., McCann, J., and Yamasaki, E. (1975), "Methods for detecting carcinogens and mutagens with the salmonella/mammalian-microsome mutagenicity test," *Mutation Research*, 31, 347–364.

Barlow, R. E., Bartholomew, D. J., Bremner, M. J., and Brunk, H. D. (1972), *Statistical Inference under Order Restriction*, New York: John Wiley & Sons.

Bornkamp, B., Pinheiro, J. C., and Bretz, F. (2009), "MCPMod - An R Package for the Design and Analysis of Dose-Finding Studies," *Journal of Statistical Software*, 29(7), 1–23.

Bretz, F. (1999), "Powerful Modification on Williams' Test on Trend," Ph.D. thesis, Universität Hannover.

— (2006), "An Extension of the Williams Trend Test to General Unbalanced Linear Models." *Computational Statistics and Data Analysis*, 50, 1735–1748.

Bretz, F. and Hothorn, L. A. (2003), "Statistical Analysis of Monotone or Non-monotone Dose-Response Data from *In Vitro* Toxicological Assays," *Alternatives to Lab Animals*, 31, Supplement 1, 81–96.

Claeskens, G. and Hjort, N. L. (2008), *Model Selection and Model Averaging*, Cambridge University Press.

Dellaportas, P., Forster, J. J., and Ntzouras, I. (2002), "On Bayesian model and variable selection using MCMC." *Statistics and Computing*, 12, 27–36.

Dunson, D. B. and Neelon, B. (2003), "Bayesian inference on order constrained parameters in generalized linear models." *Biometrics*, 59, 286–295.

Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992), "Bayesian Analysis of constrained parameter and truncated data problems using Gibbs sampling." *Journal of the American Statistical Association*, 87, 523–532.

George, E. I. and McCulloch, R. E. (1993), "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88, 881–889.

Hothorn, L. A. (2006), "Multiple comparisons and multiple contrasts in randomized dose-response trials–confidence interval oriented approaches." *Journal of Biopharmaceutical Statistics*, 16(5), 711–731.

Hothorn, T., Bretz, F., and Westfall, P. (2008), "Simultaneous Inference in General Parametric Models," *Biometrical Journal*, 50(3), 346–363.

Klugkist, I. and Hoijtink, H. (2007), "The Bayes Factor for Inequality and About Equality Constrained Models," *Computational Statistics and Data Analysis*, 51, 6367–6379.

Klugkist, I. and Mulder, J. (2008), *Bayesian Evaluation of Informative Hypotheses*, New York: Springer, chap. Bayesian estimation for inequality constrained analysis of variance, pp. 27–52.

Kuo, L. and Mallick, B. (1998), "Variable selection for regression models," *The Indian Journal of Statistics*, 60, 65–81.

Lin, D., Hothorn, L. A., Djira, G. D., and Bretz, F. (2012), *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R*, Springer, chap. Multiple Contrast Tests for Testing Dose-response Relationships Under Order Restricted Alternatives, pp. 233–247.

Marcus, R. (1976), "The Powers of Some Tests of the Equality of Normal Means against an Ordered Alternative," *Biometrika*, 63, 177–183.

Mukerjee, H., Robertson, T., and Wright, F. T. (1987), "Comparison of Several Treatments with a Control Using Multiple Contrasts," *Journal of the American Statistical Association*, 82 (399), 902–910.

Newton, M., Wang, P., and Kendziorski, C. (2007), *Bayesian Inference for gene expression and proteomics*, Cambridge university press, chap. Hierarchical mixture models for expression profiles, pp. 40–52.

Ntzoufras, I. (2002), "Gibbs Variable Selection using BUGS," *Journal of Statistical Software*, 7 (7), 1–19.

O'Hara, R. B. and Sillanpää, M. J. (2009), "Review of Bayesian variable selection methods: what, how and which." *Bayesian Analysis*, 4, 85–118.

Peddada, S., Harris, S., and Harvey, E. (2005), "ORIOGEN: Order Restricted Inference for Ordered Gene Expression Data," *Bioinformatics*, 21(20), 3933–3934.

Pinheiro, J., Bretz, F., and Branson, M. (2006), *Dose finding in drug development*, Springer, New York, chap. Analysis of dose-response studies: Modeling approaches, pp. 146–171.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference.*, John Wiley & Sons Ltd.

Whitney, M. and Ryan, L. (2009), *Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity*, John Wiley & Sons, Inc., chap. Quantifying Dose-Response Uncertainty Using Bayesian Model Averaging, pp. 165–179.

Williams, D. A. (1971), "A Test for Differences between Treatment Means When Several Dose Levels Are Compared with a Zero Dose Control," *Biometrics*, 27, 103–117.