ELSEVIER

Journal of Statistical Planning and Inference 111 (2003) 165–180

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Bayesian variable and link determination for generalised linear models

Ioannis Ntzoufras[a], Petros Dellaportas[b,*], Jonathan J. Forster[c]

[a] *Department of Business Administration, University of the Aegean, Greece*
[b] *Department of Statistics, Athens University of Economics and Business, 10434 Athens, Greece*
[c] *Department of Mathematics, University of Southampton, UK*

Abstract

In this paper, we describe full Bayesian inference for generalised linear models where uncertainty exists about the structure of the linear predictor, the linear parameters and the link function. Choice of suitable prior distributions is discussed in detail and we propose an efficient reversible jump Markov chain Monte-Carlo algorithm for calculating posterior summaries. We illustrate our method with two data examples.

Keywords: Logistic regression; Markov chain Monte-Carlo; Reversible jump

1. Introduction

Generalised linear models (GLMs) are frequently used to model the dependence of a response variable Y on a set of possible explanatory variables (predictors or covariates) $X_1, X_2, \ldots, X_p$. Briefly, a GLM assumes that the mean μ of an observation of Y is related to the explanatory variables through the linear predictor $\eta = g(\mu)$ where $g(\cdot)$ is the link function. We describe a GLM, m, by the pair (γ, S) where $\gamma \in \{0,1\}^p$ is a vector of indicator variables denoting which explanatory variables are present in the linear predictor, and S denotes other structural properties such as the response distribution, link and variance functions. Therefore, $m \in \{0,1\}^p \times \mathcal{S}$, where $\mathcal{S}$ is the set of all structural properties. Note that here a variable may be a term in a factorial model, and hence be of dimension greater than one.

* Corresponding author. Tel.: +30-1-820-3567; fax: +30-1-820-3567.
E-mail addresses: ntzoufras@aegean.gr (I. Ntzoufras), petros@aueb.gr (P. Dellaportas), jf@maths.soton.ac.uk (J.J. Forster).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00298-7

In this paper, we focus on situations where uncertainty concerns γ and the link function only, with all other structural properties of the model assumed known. Therefore, we express the possible set of models as $\mathscr{M} = \{0,1\}^p \times \mathscr{L}$, where $\mathscr{L}$ is the set of all links for the distribution under consideration. Therefore $m = (\gamma, L)$ where L is the link for model m and we denote the corresponding function of μ by $g_L(\mu)$. We also denote the n observations of the response variable by $\boldsymbol{y} = (y_1, \ldots y_n)$.

Our approach is Bayesian. For each model m, a prior distribution $f(\boldsymbol{\beta}_m|m)$ is specified for the model parameters $\boldsymbol{\beta}_m$. Models are compared using posterior model probabilities $f(m|\boldsymbol{y})$ which are calculated using

$$f(m|\boldsymbol{y}) = \frac{f(m)f(\boldsymbol{y}|m)}{\sum_{m\in\mathscr{M}} f(m)f(\boldsymbol{y}|m)},$$

where

$$f(\boldsymbol{y}|m) = \int f(\boldsymbol{y}|\boldsymbol{\beta}_m, m) f(\boldsymbol{\beta}_m|m)\, \mathrm{d}\boldsymbol{\beta}_m \tag{1}$$

is the marginal likelihood, and $f(m)$ the prior probability, for model m.

As $m = (\gamma, L)$, model uncertainty concerns two aspects; the necessary predictors of the response, and the appropriate link function. Bayesian methods for variable selection for linear and generalised linear models has recently been an area of active research. See, for example, George and McCulloch (1997), Raftery (1996) and Carlin and Chib (1995). For an overview, see Dellaportas et al. (2000, 2002). Here, we focus on link functions, although variable selection remains an integral part of our approach. Our examples concern binomial data, and we consider $\mathscr{L} = \{\text{logit}, \text{probit}, \text{log–log}, \text{complementary log} - \text{log}\}$ in Section 4.1 while more general link families are considered in Section 4.2.

General link functions can be adopted by considering a family of link functions indexed by one or more (unknown) continuous-valued parameters. For a binomial response, these approaches are often based on an inverse distribution function $F^{-1}(\mu;\theta)$ and then the continuous-valued parameter(s) θ give rise to a range of possible link functions. For example, Mallick and Gelfand (1994) considered mixtures of beta distributions while Basu and Mukhopadhyay (2000) proposed normal scale mixtures. Lang (1999) considers the link function

$$g_\rho(\mu) = m_1(\rho)F_{-\infty}(\mu) + m_2(\rho)F(\mu) + m_3(\rho)F_{\infty}(\mu),$$

where ρ is a mixing parameter to be estimated; and $F_{-\infty}(\mu) = 1 - \exp(-\mathrm{e}^{\mu})$ (extreme minimum value function), $F_{\infty}(\mu) = \exp(-\mathrm{e}^{\mu})$ (extreme maximum value function) and $F(\mu) = \mathrm{e}^{\mu}/(1+\mathrm{e}^{\mu})$ (logistic function). The mixing parameter ρ, is given a prior distribution. Aranda-Ordaz (1981) and Albert and Chib (1997) used the family of symmetric links given by

$$g_\rho(\mu) = \frac{2}{\rho} \times \frac{\mu^\rho - (1-\mu)^\rho}{\mu^\rho + (1-\mu)^\rho},$$

where $\rho = 0.0, 0.4, 1.0$ correspond to the logit, (approximately) probit and linear links, respectively. Aranda-Ordaz (1981) also proposed an asymmetric family

given by

$$g_\rho(\mu) = \frac{(1-\mu)^{-\rho} - 1}{\rho}.$$

Alternatively, Albert and Chib (1993) used link functions based on the inverse distribution function of the t distribution while Genter and Farewell (1985) proposed the inverse cumulative distribution function of a random variable which is $1/q$ times the logarithm a gamma random variable with shape and scale parameters both equal to q^{-2}. Other approaches include the link family based on the inverse distribution function of the logarithm of an F-distributed random variable (Prentice, 1976), the link family defined by

$$g_{\rho_1,\rho_2}(\mu) = \frac{\mu^{\rho_1-\rho_2} - 1}{\rho_1 - \rho_2} = \frac{(1-\mu)^{\rho_1+\rho_2} - 1}{\rho_1 + \rho_2}.$$

(Pregibon, 1980), the Box–Cox transformation based link family

$$g_\rho(\mu) = \frac{[\mu/(1-\mu)]^{-\rho} - 1}{\rho}$$

(Guerrero and Johnson, 1982), the generalization of the logit link suggested by Stukel (1988) and a family of robust link functions proposed by Haro-López et al. (2000).

The paper is organized into four further sections. In Section 2, we describe construction of the prior distributions for the unknown quantities, which are the model (γ, L) and the corresponding model parameters. The resulting posterior distributions cannot, in general, be calculated analytically. Therefore, in Section 3, we describe a method of Markov chain Monte-Carlo computation for simultaneous link and variable selection, based on reversible jump (Green, 1995). This method is illustrated with various examples in Section 4. Section 5 contains a brief discussion.

2. Prior distributions

We require a prior distribution to express prior uncertainty about the unknown quantities $m=(\gamma, L)$ and $\boldsymbol{\beta}_m$ (which we also write as $\boldsymbol{\beta}_{\gamma L}$). The prior distribution is constructed hierarchically as

$$f(\boldsymbol{\beta}_{\gamma L}, \gamma, L) = f(\boldsymbol{\beta}_{\gamma L}|\gamma, L) f(\gamma|L) f(L).$$

The prior distributions $f(L)$ and $f(\gamma|L)$ are discrete distributions over finite sets. For convenience we usually assume them to be discrete uniform. In particular, we take

$$f(L) = \frac{1}{|\mathscr{L}|}, \quad L \in \mathscr{L}.$$

However, certain combinations of explanatory variables may be considered a priori more plausible than others. For example, it may be sensible to rule out certain combinations ($f(\gamma|L) = 0$); see Section 4 for details. The performance of the posterior computations is not affected by the choice of $f(\gamma|L)$ or $f(L)$.

The remainder of this section is devoted to consideration of prior distributions for the model parameters $\boldsymbol{\beta}_{\boldsymbol{L}}$ for model $m=(\boldsymbol{\gamma},L)$. A possible family of prior distributions is

$$\boldsymbol{\beta}_{\gamma L}|\boldsymbol{\gamma},\ L\sim \mathrm{N}(\boldsymbol{\theta}_{\gamma L},\boldsymbol{\Sigma}_{\gamma L})\quad (\boldsymbol{\gamma},L)\in\{0,1\}^{p}\times\mathscr{L}.$$

We write $\boldsymbol{\beta}_{\gamma L}$ as $(\beta_{\gamma L0},\boldsymbol{\beta}^{*}_{\gamma L})$ where $\beta_{\gamma L0}$ is an intercept term (assumed present in all models). In situations where strong prior information does not exist, the prior mean of $(\beta_{\gamma L0},\boldsymbol{\beta}^{*}_{\gamma L})$ may be set at a neutral value of $(\theta_{\gamma L},\mathbf{0})$, so that all terms other than the intercept have zero prior mean; see also Raftery (1996) and Dellaportas and Forster (1999).

In the class of models which we are considering, there exist models with the same set of explanatory variables, but with different link functions. It seems sensible to require that the prior distributions for the parameters of any two such models are consistent with respect to the pattern of dependence of the response on the predictors. Therefore, we propose that, for every $\boldsymbol{\gamma}\in\{0,1\}^{p}$, there must be a consistency relationship between the means $\theta_{\gamma L}$ and between the variances $\boldsymbol{\Sigma}_{\gamma L}$ for all $L\in\mathscr{L}$.

This relationship is based on a Taylor series expansion of the link functions about $\mu=\mu_0$

$$\eta_L=g_L(\mu)=g_L(\mu_0)+(\mu-\mu_0)g'_L(\mu_0)+\cdots, \tag{2}$$

where η_L is the linear predictor corresponding to link L. Then, truncating this expansion after the linear term, we have, for every $L\in\mathscr{L}$,

$$\mu-\mu_0=\frac{\eta_L-g_L(\mu_0)}{g'_L(\mu_0)}.$$

Therefore, we have an approximate linear relationship between the linear predictors provided by any two link functions L_1 and L_2

$$\eta_{L_1}=\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,\eta_{L_2}+g_{L_1}(\mu_0)-\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,g_{L_2}(\mu_0). \tag{3}$$

Clearly this approximation will be most appropriate for values of μ close to μ_0. However, we also use it to obtain the required consistency relationship between the prior distributions as follows.

$$\boldsymbol{\eta}_{\gamma L}=\boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\gamma L}=\beta_{\gamma L0}\mathbf{1}+\boldsymbol{X}^{*}_{\boldsymbol{\gamma}}\boldsymbol{\beta}^{*}_{\gamma L}, \tag{4}$$

where $\boldsymbol{\eta}_{\gamma L}$ is the vector of linear predictors for all observations and $\boldsymbol{X}_{\boldsymbol{\gamma}}$ is the model matrix for model $m=(\boldsymbol{\gamma},L)$. From (3) and (4), we have

$$\beta_{\gamma L_1 0}=\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,\beta_{\gamma L_2 0}+g_{L_1}(\mu_0)-\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,g_{L_2}(\mu_0), \tag{5}$$

$$\boldsymbol{\beta}^{*}_{\gamma L_1}=\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,\boldsymbol{\beta}^{*}_{\gamma L_2}. \tag{6}$$

Therefore, from (5) and (6), the consistency relationship between the means for the intercept parameters of any two models with different links is

$$\theta_{\gamma L_1} = \frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,\theta_{\gamma L_2} + g_{L_1}(\mu_0) - \frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\,g_{L_2}(\mu_0) \tag{7}$$

and from (6) the consistency relationship between the variances for the model parameters of any two models with the same $\boldsymbol{\gamma}$ but different links is

$$\boldsymbol{\Sigma}_{\gamma L_1} = \boldsymbol{\Sigma}_{\gamma L_2}\left(\frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)}\right)^2. \tag{8}$$

We now require a prior mean $\theta_{\gamma L}$ for $\beta_{\gamma L0}$ and a prior variance $\boldsymbol{\Sigma}_{\gamma L}$ for $\boldsymbol{\beta}_{\gamma L}$ for each $\boldsymbol{\gamma} \in \{0,1\}^p$ for one (reference) link. The prior parameters for other links may be obtained using (7) and (8).

When strong prior information does not exist, we choose $\boldsymbol{\Sigma}_{\gamma L}$ to give a diffuse prior distribution. However, care is required as the marginal likelihood for m (1) is sensitive to the choice of $\boldsymbol{\Sigma}_{\gamma L}$. Following Kass and Wasserman (1995), we use a prior variance $\boldsymbol{\Sigma}_{\gamma L}$ which corresponds to unit prior information. For a generalised linear model $m = (\boldsymbol{\gamma}, L)$, the information matrix $\mathcal{I}(\boldsymbol{\beta}_{\gamma L})$ is given by

$$\mathcal{I}(\boldsymbol{\beta}_{\gamma L}) = \boldsymbol{X}_{\boldsymbol{\gamma}}^{\mathrm{T}} \boldsymbol{W}_{\gamma L} \boldsymbol{X}_{\boldsymbol{\gamma}},$$

where

$$\boldsymbol{W}_{\gamma L} = \operatorname{diag}\left(\frac{1}{V(\mu_i)\phi_i g'_L(\mu_i)^2}\right),$$

where μ_i represents the mean of the ith observation of the response, ϕ_i the corresponding scale parameter, and $V(\cdot)$ the variance function for the model.

The unit information matrix is given by $\mathcal{I}(\boldsymbol{\beta}_{\gamma L})/N$ where N is the number of units in the data. This is typically the number of observations of Y. The unit information prior variance is then

$$N\mathcal{I}(\boldsymbol{\beta}_{\gamma L})^{-1} = N(\boldsymbol{X}_{\boldsymbol{\gamma}}^{\mathrm{T}} \boldsymbol{W}_L \boldsymbol{X}_{\boldsymbol{\gamma}})^{-1}.$$

However, this matrix typically depends on the unknown parameters $\boldsymbol{\beta}_{\gamma L}$, through $V(\mu_i)$ and $g'_L(\mu_i)$. Therefore, in order to use it as a prior variance, we need to input prior estimates for $\boldsymbol{\beta}_{\gamma L}$. We propose to replace $\boldsymbol{\beta}_{\gamma L}$ by the prior mean $(\theta_{\gamma L}, \mathbf{0})$. Then $\eta_i = \theta_{\gamma L}$ and $\mu_i = g_L^{-1}(\theta_{\gamma L})$ for all i. Note that, if $\theta_{\gamma L}$ for different L are related by (7) and all link functions are approximated linearly by (2), then $g_L^{-1}(\theta_{\gamma L})$ is the same for all $L \in \mathscr{L}$. We denote this common value of μ by μ^*. Therefore, we use the unit information prior variance

$$\boldsymbol{\Sigma}_{\gamma L} = NV(\mu^*)g'_L(\mu^*)^2(\boldsymbol{X}_{\boldsymbol{\gamma}}^{\mathrm{T}} \operatorname{diag}(1/\phi_i)\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1}. \tag{9}$$

Note that $V(\cdot)$ does not depend on L, so unit information prior variance matrices are mutually consistent, as (8) is satisfied for $\mu_0 = \mu^*$. For convenience, the further approximation $\phi_i = \phi$ for all i may be used if the scale parameters are of a similar magnitude. This is particularly convenient if the predictors are orthogonal, in which

case the prior variance matrix becomes diagonal. For example, $\phi = \min \phi_i$ may be thought of as a 'lower bound' for a unit information prior variance.

3. Reversible jump Markov chain Monte-Carlo

Green (1995) proposed reversible jump Markov chain Monte-Carlo (RJMCMC) as a general MCMC approach appropriate for examples where model uncertainty exists. The general approach is extremely flexible, allowing quite general transitions in the joint space of models and model parameters.

Dellaportas et al. (2002) describe an application of reversible jump MCMC to variable selection. Within a model, the parameters are updated using a Gibbs sampler. Model updating is performed by local moves where single terms are added or deleted from the model. Following Dellaportas and Forster (1999), when a variable is deleted from a model, all remaining parameters are unchanged. When a variable is added to the model, a value for the parameter corresponding to the new variable is generated from a normal proposal distribution with parameters which do not depend on which other variables are present in the model. In the examples presented in Section 4, the proposal mean is calculated using the maximum likelihood estimate for the saturated model, and the variance using the curvature of the log-likelihood at this maximum likelihood estimate. Hence, different proposals are used for different link functions.

Here we adapt this approach to GLMs where additional uncertainty exists about the link function L. At any iteration, we may propose to change L to L', with $\boldsymbol{\gamma}$ remaining at its current value. Therefore, we require a value of $\boldsymbol{\beta}_{\gamma L'}$. As we already have a value of $\boldsymbol{\beta}_{\gamma L}$, it makes sense to use this information to generate $\boldsymbol{\beta}'_{\gamma L}$. We propose to generate $\boldsymbol{\beta}'_{\gamma L}$ from $\boldsymbol{\beta}_{\gamma L}$ deterministically, using (5) and (6). This transformation requires us to specify a suitable value of μ_0. One possibility is to use the approximate prior value, μ^*. However, as we are choosing μ_0 to facilitate the mobility of our reversible jump sampler, we should choose a value that makes the approximations (5) and (6) as accurate as possible. Therefore, a natural choice is $\mu_0 = \bar{y}$ or $\mu_0 = \sum \phi_i^{-1} y_i / \sum \phi_i^{-1}$. In our examples, we found that using this transformation allows efficient transition between links, which does not occur if the identity transformation is used.

Our reversible jump algorithm for variable and link selection can be summarised by sampling (in any order) $\boldsymbol{\beta}_{\gamma L}$, $\boldsymbol{\gamma}$ and L as follows:

(1) Generate each element of $\boldsymbol{\beta}_{\gamma L}$ from its log-concave univariate full conditional posterior distribution $f(\beta_{\gamma Li} | \boldsymbol{\beta}_{\gamma L \setminus i}, \boldsymbol{\gamma}, L, \boldsymbol{y})$. Either Metropolis–Hastings or adaptive rejection algorithms may be used; we have found that both methods have similar performance.

(2a) Generate a proposed variable $j \in \{1, \ldots, p\}$ to add or delete from the model with probability $1/p$. Therefore, we propose to change $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}'$ where $\gamma'_j = 1 - \gamma_j$ with all other components remaining the same.

(2b) If $\gamma_j = 0$ then

(i) Generate the additional parameter corresponding to variable j from a proposal density $q_j(\boldsymbol{\beta}'_j | L)$.

(ii) Set $\boldsymbol{\beta}'_{\gamma' L} = [\boldsymbol{\beta}_{\gamma L}, \boldsymbol{\beta}'_j]$.

(iii) Accept the proposed move with probability

$$\alpha = \min\left\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}'_{\gamma' L}, \boldsymbol{\gamma}', L) f(\boldsymbol{\beta}'_{\gamma' L}|\boldsymbol{\gamma}', L) f(\boldsymbol{\gamma}', L)}{f(\boldsymbol{y}|\boldsymbol{\beta}_{\gamma L}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_{\gamma L}|\boldsymbol{\gamma}, L) f(\boldsymbol{\gamma}, L) q_j(\boldsymbol{\beta}'_j|L)}\right\}.$$

(iv) If the proposed move is accepted, update $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\gamma L}$ by $\boldsymbol{\gamma}'$ and $\boldsymbol{\beta}'_{\gamma' L}$, respectively; otherwise leave $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\gamma L}$ unchanged.

(2c) If $\gamma_j = 1$ then

(i) Set $\boldsymbol{\beta}'_{\gamma' L}$ equal to the corresponding parameters of $\boldsymbol{\beta}_{\gamma, L}$.

(ii) Accept the proposed move with probability

$$\alpha = \min\left\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}'_{\gamma' L}, \boldsymbol{\gamma}', L) f(\boldsymbol{\beta}'_{\gamma' L}|\boldsymbol{\gamma}', L) f(\boldsymbol{\gamma}', L) q_j(\boldsymbol{\beta}_j|L)}{f(\boldsymbol{y}|\boldsymbol{\beta}_{\gamma L}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_{\gamma L}|\boldsymbol{\gamma}, L) f(\boldsymbol{\gamma}, L)}\right\}.$$

(iii) If the proposed move is accepted, update $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\gamma L}$ by $\boldsymbol{\gamma}'$ and $\boldsymbol{\beta}'_{\gamma' L}$, respectively; otherwise leave $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\gamma L}$ unchanged.

(3) (i) Propose a new link $L' \neq L$ with probability $j(L, L') = 1/(|\mathscr{L}| - 1)$.

(ii) Calculate $\boldsymbol{\beta}'_{\gamma L'}$ using (5) and (6).

(iii) Accept the proposed move with probability

$$\alpha = \min\left\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}'_{\gamma L'}, \boldsymbol{\gamma}, L') f(\boldsymbol{\beta}'_{\gamma L'}|\boldsymbol{\gamma}, L') f(\boldsymbol{\gamma}, L') j(L', L)}{f(\boldsymbol{y}|\boldsymbol{\beta}_{\gamma L}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_{\gamma L}|\boldsymbol{\gamma}, L) f(\boldsymbol{\gamma}, L) j(L, L')} \left|\frac{\partial \boldsymbol{\beta}'_{\gamma L'}}{\partial \boldsymbol{\beta}_{\gamma L}}\right|\right\}, \quad (10)$$

where

$$\left|\frac{\partial \boldsymbol{\beta}'_{\gamma L'}}{\partial \boldsymbol{\beta}_{\gamma L}}\right| = \left(\frac{g'_{L'}(\mu_0)}{g'_L(\mu_0)}\right)^{d(\boldsymbol{\gamma})}$$

and $d(\boldsymbol{\gamma})$ is the dimension (number of parameters) of the model $m = (\boldsymbol{\gamma}, L)$.

(iv) If the proposed move is accepted, update L and $\boldsymbol{\beta}_{\gamma L}$ to L' and $\boldsymbol{\beta}'_{\gamma L'}$, respectively; otherwise leave L and $\boldsymbol{\beta}_{\gamma L}$ unchanged.

4. Implementation for binomial data

4.1. Common link functions

Our examples concern observations of n binomial random variables, $Y_1, \ldots, Y_n$ (expressed as proportions) with corresponding denominators $m_1, \ldots, m_n$. The mean parameters are the binomial probabilities and $V(\mu) = \mu(1 - \mu)$. We consider the set of link functions $\mathscr{L} = \{\text{logit}, \text{probit}, \log - \log, \text{complementary } \log - \log\}$. See Table 1 for details.

We use a normal prior distribution for the model parameters for all models, as proposed in Section 2. In the absence of strong prior information, for the canonical (logistic) link function we propose a value of $\theta_{\gamma L} = 0$, and hence $\mu^* = 0.5$. The value of $\theta_{\gamma L}$ for all other links follows from (7). The unit information prior variance for

Table 1
Binomial link functions

Link	$g(\mu)$	$g'(\mu)$
Logit	$\mathrm{Log}[\mu/(1-\mu)]$	$[\mu(1-\mu)]^{-1}$
Probit	$\Phi^{-1}(\mu)$	$[\phi(\Phi^{-1}(\mu))]^{-1}$
Log–log	$-\mathrm{Log}[-\log(\mu)]$	$-[\mu\log(\mu)]^{-1}$
Complementary log–log	$\mathrm{Log}[-\log(1-\mu)]$	$-[(1-\mu)\log(1-\mu)]^{-1}$

binomial data and logistic link is, from (9),

$$\boldsymbol{\Sigma}_{\gamma L} = 4\phi \sum_{i=1}^{n} m_i (\boldsymbol{X}_{\gamma}^{\mathrm{T}} \boldsymbol{X}_{\gamma})^{-1}. \tag{11}$$

Taking $\mu^* = 0.5$, and N, the total number of units in the data equal to $\sum_{i=1}^{n} m_i$. Here, we have also approximated all $\phi_i \equiv 1/m_i$ by the common value ϕ. In the examples, we used $\phi^{-1} = \bar{m}$ and $\phi^{-1} = \max m_i$. However, the results are similar, so we only report those for $\phi^{-1} = \max(m_i)$. Note that when all m_i are equal, these coincide, and the approximation is exact. The unit prior information matrix for all other links follows from (8).

To calculate posterior summaries, we used the reversible jump algorithm described in Section 3, with $\mu_0 = \sum m_i y_i / \sum m_i$ for link transitions. For Example 2, we used orthogonal polynomials as explanatory variables, as then parameters have a similar interpretation across models. This increases mobility of transitions in $\boldsymbol{\gamma}$, without affecting the interpretation of the polynomial regression models considered. In Example 1, the dummy explanatory variables are orthogonal by construction. We do not allow arbitrary combinations of terms in the linear predictor, restricting models to those that satisfy the usual marginality restrictions.

The subsamples containing the model parameters β_j for those models with posterior probability higher than 0.05 were checked for MCMC convergence, by using the diagnostics of Geweke (1992) and Heidelberger and Welch (1983). MCMC standard errors for the model probabilities were calculated by dividing the MCMC output into 40 batches. The length of the Markov chain was determined by imposing upper bounds for these standard errors. In the results presented here we chose this upper bound to be $\max\{0.015, 0.03 f(m|\boldsymbol{y})\}$ where $f(m|\boldsymbol{y})$ is estimated using the MCMC output.

4.1.1. Example 1

We first consider a data set analysed by Healy (1988) and presented in Table 2. The data reflect the relationship between the number of survivals, the patient condition (more or less severe) and the received treatment (antitoxin medication or not). Dellaportas et al. (2000) analysed these data using a Bayesian approach incorporating uncertainty about variables but not links.

A total of 110,000 iterations of reversible jump MCMC were used, with the first 10,000 discarded as burn in. The Markov chain mixed well. In particular, mobility

Table 2
Example 1: Healy (1988)

Severity(A)	Antitoxin (B)	Death	Survivals
More severe	Yes	15	6
	No	22	4
Less severe	Yes	5	15
	No	7	5

Table 3
Posterior model probabilities for Example 1

Link	Linear predictor				
	1	1+B	1+A	1+A+B	1+A+B+AB
Logit	0.001	0.002	0.108	0.146	0.028
Probit	0.001	0.002	0.098	0.121	0.021
Log–log	0.001	0.002	0.097	0.088	0.021
Complementary log–log	0.001	0.003	0.097	0.141	0.023

between different link functions was extremely high, with a link change occurring on average in 77% of iterations.

The posterior model probabilities are presented in Table 3. Of the 20 models considered, only those with linear predictors $1+A$ and $1+A+B$ have substantial posterior probability. For the model $1+A$ the posterior probabilities are close for all four links. This is to be expected as this model simply implies that, regardless of the link function, the data are modelled by two binomial probabilities, one for each level of severity. The only difference is in the prior distributions, and this is minimised by the consistency relationships discussed in Section 2. For the model $1+A+B$, a greater difference between link functions is evident, but this difference is still small, as the binomial probabilities are not close to the extremes (0,1) where the difference between link functions is greatest.

The similarity between the model probabilities suggests that selection of a single model may not be appropriate. The model probabilities may be used as weights in a 'model averaged' posterior inference which fully accounts for model uncertainty (Draper, 1995).

The mixing of the Markov chain was satisfactory, producing Monte-Carlo standard errors for all posterior model probabilities less than 0.008. To investigate the sensitivity of the approach to the choice of μ_0, required for parameter transformations when a change of links is proposed, we performed MCMC runs using seven different values of μ_0, and also without any transformation. Selected MCMC performance statistics are presented in Table 4. Our proposed transformation ($\mu_0 = 0.401$) seems to be a good choice. For all values $0.25 \leqslant \mu_0 \leqslant 0.75$ all posterior model probabilities are estimated with a Monte-Carlo standard error less than 1.5%. Without any transformation, the reversible jump sampler performs poorly.

Table 4
Example 1: Performance of reversible jump sampler, based on different transformation parameters μ_0 when a link change is proposed.

Proposal μ_0	Link jumps accepted (%)	Maximum model MCMC s.e. (%)	Models with MCMC s.e. $> \max\{0.015, 0.03 f(m\|\boldsymbol{y})\}$
0.05	16.7	1.88	4
0.25	59.5	0.84	0
0.40	76.9	0.80	0
0.50	69.4	0.83	0
0.60	58.1	1.09	0
0.75	38.3	1.27	0
0.95	4.8	4.44	6
No transformation*	12.1	11.28	9

*log–log models were not visited by the chain

Table 5
Example 2: Dobson (1983)

Concentration (X)	Total (M)	Number killed (Y)
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	52
1.8610	62	61
1.8839	60	60

4.1.2. Example 2

The second data set we consider was analysed by Dobson (1983) and is presented in Table 5. This data set describes the number of beetles killed in each of eight groups after a 5 h exposure to carbon disulphide. Spiegelhalter et al. (1996) analysed these data using a Bayesian approach but did not explicitly incorporate uncertainty either about variables or links.

A total of 210,000 iterations of reversible jump MCMC were used. The first 10,000 were discarded as burn in, and the remaining sample was thinned by a factor of 5 due to storage constraints. The Markov chain mixed well, although not quite as well as Example 1; a link change occurred on average in 8% of iterations.

The posterior model probabilities are presented in Table 6. A much greater distinction between link functions is observed here, compared with Example 1, as several of the binomial probabilities are close to the extreme value 1. The dominant model is the simple linear model with complementary log–log link. For the other link functions, the linear model does not fit well, and the quadratic model has a substantially higher posterior probability. Our approach allows a quantifiable comparison, in terms of

Table 6
Posterior model probabilities for Example 2

Link	Linear predictor		
	$1+X$	$1+X+X^2$	$1+X+X^2+X^3$
Logit	0.018	0.072	0.008
Probit	0.026	0.058	0.005
Log–log	0.000	0.024	0.004
Complementary log–log	0.714	0.065	0.006

Table 7
Example 2: Performance of reversible jump sampler, based on different transformation parameters μ_0 when a link change is proposed.

Proposal μ_0	Link jumps accepted (%)	Maximum model MCMC s.e. (%)	Models with MCMC s.e. $> \max\{0.015, 0.03 f(m\|\boldsymbol{y})\}$
0.05*	0.05	24.34	4
0.25$^+$	5.26	9.07	3
0.40	6.29	3.45	2
0.50	7.48	2.44	1
0.60$^+$	7.94	2.04	0
0.75$^+$	6.56	2.11	1
0.95$^+$	0.23	38.92	8
No transformation*	0.14	21.54	5

*log–log models were not visited by the chain; ^+1+X log–log model was not visited by the chain

posterior probabilities, of models which are not nested, such as the linear complementary log–log model and the quadratic models with other links. The posterior model probabilities favour the more parsimonious model, as one would desire.

As in Example 1, we investigated the sensitivity of the approach to the choice of μ_0. The results are presented in Table 7. Again, the results indicate that our proposed transformation (here $\mu_0 = 0.60$) is within the range of values for which the algorithm works well, and again the reversible jump sampler performed poorly when no transformation was used.

4.2. General link families

As an alternative to the link functions presented in the previous section we may adopt more flexible link families, of the form discussed in Section 1. For illustration, we consider here the comparison of the inverse t-link family proposed by Albert and Chib (1993) and the log–gamma link family proposed by Genter and Farewell (1985).

The inverse t-link family is given by $g_{\mathcal{T}_\theta}(\mu) = F_{\mathcal{T}}^{-1}(\mu;\theta)$ where $F_{\mathcal{T}}(\cdot;\theta)$ is the distribution function of a t distribution with $\theta > 1$ degrees of freedom. This link family

includes as a special case the probit link function ($\theta \to \infty$). Furthermore, Albert and Chib (1993) argue that the t-link with $\theta = 8$ is a reasonable approximation to the logit link. The log–gamma link family is given by $g_{\Gamma_\theta}(\mu) = (1/|\theta|)\log(\theta^2 F_\Gamma^{-1}[\mu; \theta^{-2}])$ if $\theta > 0$ and $g_{\Gamma_\theta}(\mu) = (1/|\theta|)\log(\theta^2 F_\Gamma^{-1}[1-\mu; \theta^{-2}])$ if $\theta < 0$, where $F_\Gamma(\cdot;\alpha)$ is the distribution function of gamma distribution with mean and variance equal to α. This family includes as special cases the probit link ($\theta \to 0$), the log–log link ($\theta = -1$) and the complementary log–log link ($\theta = 1$). The link parameter θ controls the symmetry of the link function while the t-link is symmetric for all values of θ. Hence, the link function can be (at least approximately) any one of the four links we considered in Section 4.1, as well as a wide range of alternative symmetric and asymmetric functions.

For the link indicator L, we now substitute the pair (ℓ, θ) where ℓ is a binary link family indicator taking value 0 for t-link and 1 for log–gamma link. Note that the interpretation of θ depends strongly on ℓ. The hierarchy of the prior distribution is slightly changed to

$$f(\boldsymbol{\beta}_{\gamma L}, \boldsymbol{\gamma}, \ell, \theta) = f(\boldsymbol{\beta}_{\gamma L}|\boldsymbol{\gamma}, \ell, \theta) f(\theta|\ell) f(\boldsymbol{\gamma}, \ell).$$

The prior for the model coefficients $\boldsymbol{\beta}_{\gamma L}$ is chosen to be a multivariate normal distribution with zero mean and covariance matrix given by (8) and (11). The derivatives involved in (8) are given by

$$g'_{\mathscr{T}_\theta}(\mu_0) = (f_{\mathscr{T}}[F_{\mathscr{T}}^{-1}(\mu_0; \theta); \theta])^{-1} \tag{12}$$

and

$$g'_{\Gamma_\theta}(\mu_0) = \begin{cases} (|\theta| F_\Gamma^{-1}(\mu_0; \theta^{-2}) f_\Gamma[F_\Gamma^{-1}(\mu_0; \theta^{-2}); \theta^{-2}])^{-1} & \text{if } \theta > 0, \\ (|\theta| F_\Gamma^{-1}(1-\mu_0; \theta^{-2}) f_\Gamma[F_\Gamma^{-1}(1-\mu_0; \theta^{-2}); \theta^{-2}])^{-1} & \text{if } \theta < 0. \end{cases} \tag{13}$$

For the link parameter θ, we specify a prior which reflects the fact that the shape of the link function changes most rapidly with respect to θ when θ is close to 1 for the t-link and close to 0 for the log–gamma link. Hence, we choose priors which have highest density at these values, with tails monotonically decreasing, but still quite heavy, to reflect genuine uncertainty about θ. Therefore, for the t-link, we choose $f(\theta|\ell = 0) = \theta^{-2}$ (Pareto) and for the log–gamma link, the prior distribution for $\theta|\ell = 1$ is t_3. Finally, we specify $f(\boldsymbol{\gamma}, \ell) = (2|\mathscr{M}|)^{-1}$ where $\mathscr{M}$ here represents the parameter space for $\boldsymbol{\gamma}$.

For the implementation of the reversible jump algorithm, we divide step 3 of the procedure proposed in Section 3 into two sub-steps: one (3a) for updating the link family parameter θ and the other (3b) for updating the indicator ℓ. The two steps are as follows:

(3a) (i) If $\ell = 0$ propose θ' from a uniform distribution over $(\max\{1, \theta - c_0/2\}, \theta + c_0/2)$, and if $\ell = 1$ propose θ' from an $N(\theta, c_1^2)$ proposal distribution where c_0 and c_1 are tuning parameters.

(ii) Calculate $\boldsymbol{\beta}'_{\gamma L'}$ using (5), (6) and (12) or (13).

(iii) Accept the proposed move with probability (10) in which the Jacobian is replaced by

$$\left|\frac{\partial \boldsymbol{\beta}'_{\gamma L'}}{\partial \boldsymbol{\beta}_{\gamma L}}\right| = \left[\left(\frac{g'_{\mathcal{T}_{\theta'}}(\mu_0)}{g'_{\mathcal{T}_{\theta}}(\mu_0)}\right)^{1-\ell}\left(\frac{g'_{\Gamma_{\theta'}}(\mu_0)}{g'_{\Gamma_{\theta}}(\mu_0)}\right)^{\ell}\right]^{d(\gamma)},$$

where $g'_{\mathcal{T}}$ and g'_{Γ} are given by (12) and (13), and $j(L',L)/j(L,L') = (\theta + c/2 - \max\{1, \theta - c/2\})/(\theta' + c/2 - \max\{1, \theta' - c/2\})$ if $\ell = 0$, and 1 otherwise.

(iv) If the proposed move is accepted then update θ by θ' and $\boldsymbol{\beta}_{\gamma L}$ by $\boldsymbol{\beta}'_{\gamma L'}$; otherwise leave θ and $\boldsymbol{\beta}_{\gamma L}$ unchanged.

(3b) (i) If $\ell = 0$, propose $\ell' = 1$ and generate candidate θ' from an $N(\bar{\theta}_1, S_1^2)$ distribution, where $\bar{\theta}_1$ and S_1^2 are posterior estimates based on a pilot reversible jump MCMC run, with ℓ fixed at 1. If $\ell = 1$, propose $\ell' = 0$ and generate candidate θ' as $1 + \exp(u)$ where u has an $N(\bar{\theta}_0, S_0^2)$ distribution, and $\bar{\theta}_0$ and S_0^2 are posterior estimates based on a pilot reversible jump MCMC run, with ℓ fixed at 0. Alternatively, using the prior as a proposal distribution is possible.

(ii) Calculate $\boldsymbol{\beta}'_{\gamma L'}$ by transforming $\boldsymbol{\beta}_{\gamma L}$ using (5), (6), (12) and (13).

(iii) Accept the proposed move with probability (10) in which the Jacobian is replaced by

$$\left|\frac{\partial \boldsymbol{\beta}'_{\gamma L'}}{\partial \boldsymbol{\beta}_{\gamma L}}\right| = \left[\left(\frac{g'_{\Gamma_{\theta'}}(\mu_0)}{g'_{\mathcal{T}_{\theta}}(\mu_0)}\right)^{1-\ell}\left(\frac{g'_{\mathcal{T}_{\theta'}}(\mu_0)}{g'_{\Gamma_{\theta}}(\mu_0)}\right)^{\ell}\right]^{d(\boldsymbol{\gamma})},$$

where $g'_{\mathcal{T}}$ and g'_{Γ} are given by (12) and (13), and $j(L',L)/j(L,L')$ is based on the proposal densities for θ' implied by (i) above.

(iv) If the proposed move is accepted then update L and $\boldsymbol{\beta}_{\gamma L}$ by L' and $\boldsymbol{\beta}'_{\gamma L'}$ respectively, otherwise leave them unchanged.

4.2.1. Examples 1 and 2 revisited

For Example 1, the MCMC output was based on 10,000 burn-in and a further 200,000 iterations. The mixing was reasonably satisfactory, with a link family change occurring on average in 55% of iterations, while the link parameter θ was updated using step 3a in over 40% of iterations for both $\ell = 0$ and 1.

The posterior model probabilities are presented in Table 8. There is some support for a symmetric link function, with the t-link being generally preferred, but there remains a large amount of uncertainty about the link (degrees of freedom) parameter. Indeed, for no model is the posterior for θ significantly different from the prior. Compared with Table 6, slightly more weight is now given to models which include both A and B main effects. Again, given the wide range of models which are supported, a model-averaged inference is attractive here. All posterior model probabilities were estimated with Monte-Carlo standard errors less than 1.5%.

For Example 2, we used a total of 400,000 iterations of reversible jump MCMC (after a burn-in of 10,000), saving every 10th observation due to storage constraints.

Table 8
Example 1: Summary of the posterior distribution $f(\gamma,\ell,\theta|\boldsymbol{y})$ for models with $f(\gamma,\ell|\boldsymbol{y})>0.01$

γ	$\ell=0$ (Student link)				$\ell=1$ (Log–gamma link)			
	$f(\gamma,\ell\|\boldsymbol{y})$	Quantiles of $f(\theta\|\gamma,\ell,\boldsymbol{y})$			$f(\gamma,\ell\|\boldsymbol{y})$	Quantiles of $f(\theta\|\gamma,\ell,\boldsymbol{y})$		
		2.5%	50%	97.5%		2.5%	50%	97.5%
$1+A+B$	0.36	1.02	1.54	18.68	0.17	−2.26	0.23	2.96
$1+A$	0.20	1.02	1.72	25.88	0.14	−3.26	−0.03	2.73
$1+A+B+AB$	0.09	1.01	1.43	14.45	0.03	−3.17	0.03	2.90
(prior)		1.03	2	40		−3.18	0	3.18

Table 9
Example 2: Summary of the posterior distribution $f(\gamma,\ell,\theta|\boldsymbol{y})$

γ	$\ell=0$ (Student link)				$\ell=1$ (Log–gamma link)			
	$f(\gamma,\ell\|\boldsymbol{y})$	Quantiles of $f(\theta\|\gamma,\ell,\boldsymbol{y})$			$f(\gamma,\ell\|\boldsymbol{y})$	Quantiles of $f(\theta\|\gamma,\ell,\boldsymbol{y})$		
		2.5%	50%	97.5%		2.5%	50%	97.5%
$1+X$	0.03	1.31	5.23	106.29	0.60	0.15	0.90	1.74
$1+X+X^2$	0.15	1.26	3.74	61.59	0.17	−1.06	0.32	2.21
$1+X+X^2+X^3$	0.03	1.11	2.56	30.70	0.02	−1.49	0.27	2.10
(prior)		1.03	2	40		−3.18	0	3.18

A change of link family occurred only on 8.4% of iterations while link parameter θ was updated in around 40% of iterations. The posterior model probabilities are presented in Table 9 and resemble the results presented in Table 6, in that for these data, a simple linear model with asymmetric link function is supported. Hence, the log–gamma link is generally preferred to the t-link. For the preferred linear predictor $(1+X)$ the posterior median for the log–gamma link parameter is 0.90, which is close to the special case of the complementary log–log link function, the preferred link in Table 6. With a larger range of observed proportions, these data provide more information about the link function and larger discrepancies between the prior and posteriors for θ are evident. Nevertheless, for linear predictors other than $1+X$ there seems to be much less information concerning θ. This is compatible with our previous results in Table 6. Marginal probabilities for the three linear predictors are also similar to those in Table 6. Again, all posterior model probabilities have been estimated with Monte-Carlo standard error less than 1.5%.

As before, in both examples, we tested the efficiency of our suggested methodology by rerunning the chain without any parameter transformation on change of L. Again, the transformation proved extremely useful in assisting mobility of the chain. For both examples, changes in ℓ were accepted over twice as often with the transformation, and changes in θ for the log–gamma link were also much more frequent. The transformation did not have such a large effect on the acceptance of proposed changes in θ for the t-link. Without the transformation, standard errors for several models exceeded our threshold of 1.5%.

5. Discussion

In this paper, we have presented full Bayesian inference for generalised linear models where uncertainty exists about the structure of the linear predictor, the linear parameters and the link function.

Dellaportas et al. (2002) described an alternative MCMC approach under model uncertainty based on an 'independence sampler', a special case of the reversible jump formulation. They described connections between this sampler, and the MCMC model selection approach proposed by Carlin and Chib (1995), based on the Gibbs sampler. The independence sampler requires a full set of model parameters ($\boldsymbol{\beta}_{\gamma L}$) to be proposed whenever a new model $(\boldsymbol{\gamma}, L)$ is proposed. Deterministic transformations of the type allowed by general RJMCMC are not permitted. Our experience with the independence sampler is that it can be made to work efficiently, but that in order to achieve this, it is usually necessary to perform a greater amount of prior tuning than is required for the kind of RJMCMC approach described in the current paper. Further details and comparisons are given by Dellaportas et al. (2002). In general, an independence sampler performs best when the proposal closely resembles the posterior, and in complex examples this is difficult to achieve without a large amount of pilot MCMC exploration.

The linear predictor and the link function are two components of a generalised linear model. We have assumed that the other aspects of the model; the distribution, associated variance function, and any scale parameters, are known. In principle uncertainty about these aspects could also be incorporated in a full Bayesian analysis, and this is an area of ongoing research.

References

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Amer. Statist. Assoc. 88, 669–679.

Albert, J.H., Chib, S., 1997. Bayesian tests and model diagnostics in conditionally independent hierarchical models. J. Amer. Statist. Assoc. 92, 916–925.

Aranda-Ordaz, F.J., 1981. On two families of transformations to additivity for binary response data. Biometrika 68, 357–363.

Basu, S., Mukhopadhyay, S., 2000. Binary response regression with normal scale mixture links. In: Dey, D.K., Gosh, S., Mallick, B. (Eds.), Generalised Linear Models: A Bayesian Perspective. Marcel Dekker, New York, pp. 231–241.

Carlin, B.P., Chib, S., 1995. Bayesian model choice via Markov chain Monte-Carlo methods. J. Roy. Statist. Soc. 57, 473–484.

Dellaportas, P., Forster, J.J., 1999. Markov chain Monte-Carlo model determination for hierarchical and graphical log-linear models. Biometrika 86, 615–633.

Dellaportas, P., Forster, J.J., Ntzoufras, I., 2000. Bayesian variable selection using the Gibbs sampler. In: Dey, D.K., Ghosh, S., Mallick, B. (Eds.), Generalised Linear Models: A Bayesian Perspective. Marcel Dekker, New York, pp. 271–286.

Dellaportas, P., Forster, J.J., Ntzoufras, I., 2002. On Bayesian model and variable selection using MCMC. Statistics and Computing 12, 27–36.

Dobson, A.J., 1983. An Introduction to Statistical Modelling. Chapman and Hall, London.

Draper, D., 1995. Assessment and propagation of model uncertainty (with discussion). J. Roy. Statist. Soc. 57, 45–98.

Genter, F.C., Farewell, V.T., 1985. Goodness-of-link testing in ordinal regression models. Canad. J. Statist. 13, 37–44.

George, E.I., McCulloch, R.E., 1997. Approaches to Bayesian variable selection. Stat. Sinica 7, 339–373.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 4. Oxford University Press, Oxford, pp. 169–193.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Guerrero, V.M., Johnson, R.A., 1982. Use of the Box–Coxtransformation with binary response models. Biometrika 69, 309–314.

Haro-López, R.A., Mallick, B.K., Smith, A.F.M., 2000. Binary regression using data adaptive robust link functions. In: Dey, D.K., Ghosh, S., Mallick, B. (Eds.), Generalised Linear Models: A Bayesian perspective. Marcel Dekker, New York, pp. 243–253.

Healy, M.J.R., 1988. Glim: An Introduction. Oxford University Press, Oxford.

Heidelberger, P., Welch, P.D., 1983. Simulation run length control in the presence of an initial transient. Oper. Res. 31, 1109–1144.

Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. J. Amer. Statist. Assoc. 90, 928–934.

Lang, J.B., 1999. Bayesian ordinal and binary regression models with a parametric family of mixture links. J. Comp. Statist. Data Anal. 31, 59–87.

Mallick, B.K., Gelfand, A.E., 1994. Generalized linear models with unknown number of components. Biometrika 81, 237–245.

Pregibon, D., 1980. Goodness of link tests for generalized linear models. Appl. Statist. 29, 15–24.

Prentice, R.L., 1976. Generalization of the probit and logit methods for dose response curves. Biometrics 32, 761–768.

Raftery, A.E., 1996. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. Biometrika 83, 251–266.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. BUGS 0.5: Examples, Vol. 2. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Stukel, T.A., 1988. Generalized logistic models. J. Amer. Statist. Assoc. 83, 426–431.