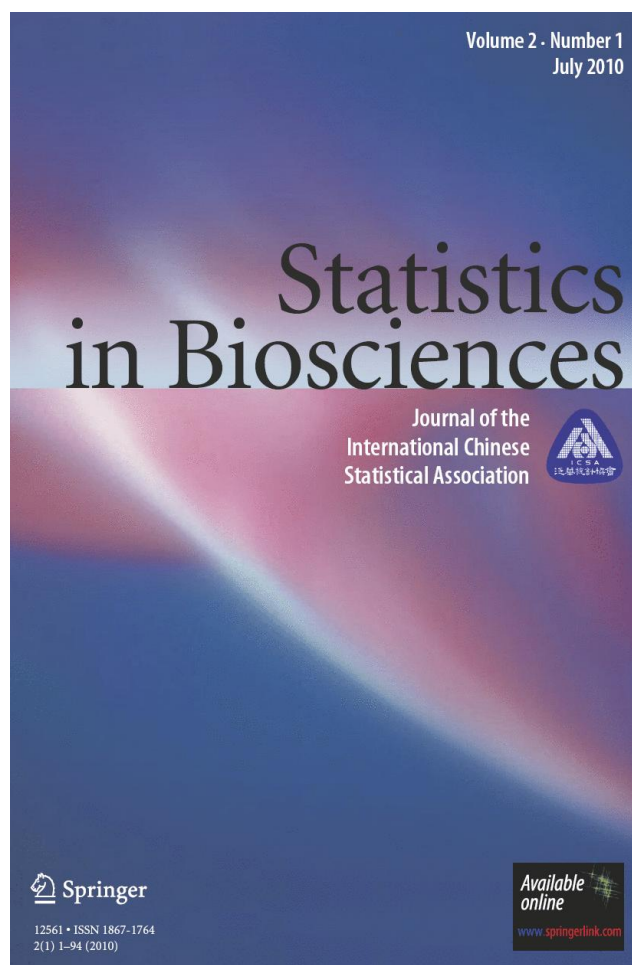


ISSN 1867-1764, Volume 2, Number 1



**This article was published in the above mentioned Springer issue.
The material, including all portions thereof, is protected by copyright;
all rights are held exclusively by Springer Science + Business Media.
The material is for personal use only;
commercial use is not permitted.
Unauthorized reproduction, transfer and/or use
may be a violation of criminal as well as civil law.**

Evaluating the Impact of Prior Assumptions in Bayesian Biostatistics

Satoshi Morita · Peter F. Thall · Peter Müller

Received: 12 September 2009 / Accepted: 5 March 2010 / Published online: 24 March 2010
© International Chinese Statistical Association 2010

Abstract A common concern in Bayesian data analysis is that an inappropriately informative prior may unduly influence posterior inferences. In the context of Bayesian clinical trial design, well chosen priors are important to ensure that posterior-based decision rules have good frequentist properties. However, it is difficult to quantify prior information in all but the most stylized models. This issue may be addressed by quantifying the prior information in terms of a number of hypothetical patients, i.e., a prior effective sample size (ESS). Prior ESS provides a useful tool for understanding the impact of prior assumptions. For example, the prior ESS may be used to guide calibration of prior variances and other hyperprior parameters. In this paper, we discuss such prior sensitivity analyses by using a recently proposed method to compute a prior ESS. We apply this in several typical settings of Bayesian biomedical data analysis and clinical trial design. The data analyses include cross-tabulated counts, multiple correlated diagnostic tests, and ordinal outcomes using a proportional-odds model. The study designs include a phase I trial with late-onset toxicities, a phase II trial that monitors event times, and a phase I/II trial with dose-finding based on efficacy and toxicity.

S. Morita (✉)

Department of Biostatistics and Epidemiology, Yokohama City University Medical Center,
4-57 Urafune-cho, Minami-ku, Yokohama 232-0024, Japan
e-mail: smorita@urahp.yokohama-cu.ac.jp

P.F. Thall · P. Müller

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX,
USA

P.F. Thall

e-mail: rex@mdanderson.org

P. Müller

e-mail: pmueller@mdanderson.org

Keywords Bayesian biostatistics · Bayesian clinical trial design · Bayesian analysis · Effective sample size · Parametric prior distribution

1 Introduction

Understanding the strength of prior assumptions relative to the likelihood is a fundamental issue when applying Bayesian methods. The processes of formulating a putatively non-informative prior or eliciting a prior from an expert area typically require one to make many arbitrary choices, including the choice of particular distributional forms and numerical hyperparameter values. In practice, these choices are often dictated by technical convenience. A common criticism of Bayesian analysis is that an inappropriately informative prior may unduly influence posterior inferences and decisions. However, it is difficult to quantify and critique prior information in all but the most stylized models. These concerns may be addressed by quantifying the prior information in terms of an equivalent number of hypothetical patients, i.e., a prior effective sample size (ESS). Such a summary allows one to judge the relative contributions of the prior and the data to the final conclusions. A useful property of prior ESS is that it is readily interpretable by any scientifically literate reviewer without requiring expert mathematical training. This is important, for example, for consumers of clinical trial results.

The purpose of this paper is to discuss prior sensitivity analyses in Bayesian biostatistics by computing the prior ESS for six case studies chosen from the recent literature. We apply an ESS method proposed by Morita, Thall and Müller (MTM) [7]. Some of our case studies require prior ESS values for a subvector θ_1 of the parameter vector $\theta = (\theta_1, \theta_2)$. The general definition allows the ESS to be specified for a subvector θ_1 of $\theta = (\theta_1, \theta_2)$. However, $\text{ESS}(\theta_1) + \text{ESS}(\theta_2)$ typically does not equal $\text{ESS}(\theta_1, \theta_2)$, because θ_1 in its marginal distribution often has a very different meaning than θ_1 in the joint distribution of (θ_1, θ_2) .

The case studies consist of three Bayesian data analyses and three study designs. The data analysis examples include small-sample cross-tabulated counts from an animal experiment to evaluate mechanical ventilator devices, bivariate normal modeling of paired data from multiple correlated diagnostic serologic tests, and proportional-odds modeling of ordinal outcomes arising from a study of viral effects in chick embryos. The study design examples include a phase I trial with dose-finding using the time-to-event continual reassessment method (TITE-CRM) [2], a phase II trial with a stopping rule for monitoring event times, and a phase I/II clinical trial in which doses were assigned based on both efficacy and toxicity.

Section 2 provides a motivating example. In Sect. 3, we briefly summarize MTM. We discuss prior sensitivity in the real examples of Bayesian data analyses and study designs in Sect. 4. We close with a brief discussion in Sect. 5.

2 A Motivating Example

The following example illustrates how the prior ESS may be used as an index of prior informativeness in a Bayesian sensitivity analysis and as a tool for critiquing a Bayesian data analysis when interpreting or formally reviewing the analysis.

Carlin [1] analyzed small-sample contingency table data from an experiment carried out to examine the effects of mechanical ventilator devices on lung damage in rabbits. In the experiment, the lungs of newborn rabbits were altered to simulate lung defects seen in human infants with underdeveloped lungs due to premature birth. The aim was to learn about the joint effects of different frequency and amplitude settings of the ventilators on lung damage. Six groups of six to eight animals each were compared using a factorial design with three frequency values crossed with two amplitudes. For amplitude g ($= 1, 2$ for 20 and 60, respectively) and frequency h ($= 1, 2, 3$ for 5, 10, and 15 Hz, respectively), let $Y_{g,h}$ denote the number of animals with lung damage out of $n_{g,h}$ studied and let $\pi_{g,h}$ denote the probability of lung damage in a cell (g, h) . The data are shown in Table 1. Each cell reports the empirical frequency $Y_{g,h}/n_{g,h}$. Carlin [1] assumed the following logistic regression model:

$$\pi_{g,h}(\boldsymbol{\theta}) = \text{logit}^{-1}(\mu + \alpha_g + \beta_h + \gamma_h I_{(g=2)}),$$

where $I_{(g=2)}$ indicates $g = 2$. Note that α_g and β_h are main effects for amplitude and frequency, while γ_h is the interaction effect at frequency h and amplitude $g = 2$. Hence, the model parameter is $\boldsymbol{\theta} = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3)$, with dimension $d = 9$. Carlin [1] assumed independent normal priors for μ , $\{\alpha_g\}$, $\{\beta_h\}$, and $\{\gamma_h\}$:

$$\mu \sim N(0, 1000^2), \quad \alpha_g \sim N(0, \tilde{\sigma}_\alpha^2), \quad \beta_h \sim N(0, \tilde{\sigma}_\beta^2), \quad \gamma_h \sim N(0, \tilde{\sigma}_\gamma^2). \quad (1)$$

Since the numbers of animals studied were very small, Carlin [1] explored the effect of a range of non-informative prior distributions in the analysis.

We use prior ESS to investigate sensitivity of the inferences to hyperparameter values by considering ten alternative choices that cover a range of reasonably non-informative settings. The ten hyperparameter choices, labeled N1 to N10, are shown in Table 2. We add the four priors, N1 to N4, which would have smaller ESS values than those considered by Carlin [1] for priors N5 to N10. We apply MTM's method to compute an overall ESS for $p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$, and also ESS_μ , ESS_α , ESS_β , and ESS_γ for the marginal priors on the subvectors μ , $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ of $\boldsymbol{\theta}$.

The prior N10 with $\tilde{\sigma}_\alpha = \tilde{\sigma}_\beta = \tilde{\sigma}_\gamma = 0.5$ has $\text{ESS} = 36.6$, so that its impact is roughly equal to that of the data (sample size $n = 38$) on the posterior inference. Thus, based on its ESS, the prior N10 may be criticized as being overly informative. Moreover, this prior has $\text{ESS}_\gamma = 96.1$, thus it assumes very high prior information for the interaction effects. Because the prior means of the interactions are 0, this prior has the effect of shrinking the posterior estimates of the interaction parameters excessively toward 0. This illustrates two important points. First, a seemingly reasonable choice of $(\tilde{\sigma}_\alpha, \tilde{\sigma}_\beta, \tilde{\sigma}_\gamma)$ may give an excessively informative prior. Second, it is important to evaluate not only the overall ESS, but also the ESS values of subvectors of $\boldsymbol{\theta}$ of particular interest. Such ESS computations help readers to interpret the reported posterior results, such as the estimates between the frequency groups displayed in Carlin (Fig. 5.1) [1]. We will revisit this example below in Sect. 4.

Table 1 Lung damage data from Carlin [1]. Each cell contains the number of animals with lung damage over the number studied at the (amplitude, frequency) combination

Amplitude	Frequency (Hz)		
	5	10	15
20	1/6	0/6	0/6
60	4/6	0/6	4/8

Table 2 Summary of priors (standard error parameters) and computed ESSs of the example for cross-tabulated counts in Carlin [1]

	$\tilde{\sigma}_\alpha$	$\tilde{\sigma}_\beta$	$\tilde{\sigma}_\gamma$	ESS	ESS $_\mu$	ESS $_\alpha$	ESS $_\beta$	ESS $_\gamma$
N1	2.0	2.0	2.0	2.3	<0.001	2.0	3.0	6.0
N2	1.5	2.0	2.0	3.7	<0.001	3.6	3.0	6.0
N3	2.0	1.5	2.0	3.0	<0.001	2.0	5.3	6.0
N4	2.0	2.0	1.5	3.0	<0.001	2.0	3.0	10.6
N5	1.5	1.5	1.5	4.1	<0.001	3.6	5.3	10.6
N6	1.5	1.5	1.0	6.0	<0.001	3.6	5.3	23.9
N7	1.5	1.5	0.5	16.3	<0.001	3.6	5.3	96.1
N8	0.5	0.5	1.5	24.4	<0.001	32.0	48.0	10.6
N9	0.5	0.5	1.0	26.3	<0.001	32.0	48.0	24.0
N10	0.5	0.5	0.5	36.6	<0.001	32.0	48.0	96.1

3 Prior Effective Sample Size

We briefly summarize the definition of ESS proposed by MTM [7]. While the discussion following this section does not require these details, we include this brief review for completeness.

Let $f(Y | \theta)$ be the sampling model for a random vector Y indexed by parameter vector $\theta = (\theta_1, \dots, \theta_d)$. We use $f(Y | \theta)$ generically to denote either a probability density function (pdf) or a probability mass function (pmf). The ESS is defined for a given prior $p(\theta | \tilde{\theta})$ on θ , having hyperparameters $\tilde{\theta}$, with respect to the sampling model $f(Y | \theta)$. The approach of MTM is constructive. First, an ε -information prior $q_0(\theta | \theta_0)$ is defined that is similar to $p(\theta | \tilde{\theta})$ but is very vague in a suitable sense. The ESS is then defined to be the sample size m of outcomes $\mathbf{Y}_m = (Y_1, \dots, Y_m)$ that, starting with $q_0(\theta | \theta_0)$, yields a posterior $q_m(\theta | \mathbf{Y}_m)$ very close to $p(\theta | \tilde{\theta})$. While the ESS can be obtained analytically in some cases, in most applications numerical methods must be used.

This constructive definition may be understood in terms of the simple example where $p(\theta | \tilde{\theta})$ is a beta distribution, $\text{Be}(\tilde{\alpha}, \tilde{\beta})$, for which the ESS is commonly considered to be $\tilde{\alpha} + \tilde{\beta}$. In this case, the ε -information prior $q_0(\theta | \theta_0)$ is specified as $\text{Be}(\tilde{\alpha}/c, \tilde{\beta}/c)$ using an arbitrarily large value $c > 0$, so that the ESS $(\tilde{\alpha} + \tilde{\beta})/c = \varepsilon$ of q_0 is very small while the mean $\tilde{\alpha}/(\tilde{\alpha} + \tilde{\beta})$ is the same as that of $p(\theta | \tilde{\theta})$. If $m = \tilde{\alpha} + \tilde{\beta}$ observations are obtained with $\tilde{\alpha}$ successes and $\tilde{\beta}$ failures then, starting with the ε -information prior $q_0(\theta | \theta_0)$, the posterior would be $\text{Be}(\tilde{\alpha} + \tilde{\alpha}/c, \tilde{\beta} + \tilde{\beta}/c)$, which has $\text{ESS} = m + \varepsilon \doteq m$.

For the general construction, denote $\mathbf{Y}_m = (Y_1, \dots, Y_m)$, with $Y_i \sim f(Y_i | \boldsymbol{\theta})$ an i.i.d. sample, so the likelihood is $f_m(\mathbf{Y}_m | \boldsymbol{\theta}) = \prod_{i=1}^m f(Y_i | \boldsymbol{\theta})$. An ε -information prior, $q_0(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0)$, is defined by requiring matching means, $E_{q_0}(\boldsymbol{\theta}) = E_p(\boldsymbol{\theta})$, and correlations, $\text{Corr}_{q_0}(\theta_j, \theta_{j'}) = \text{Corr}_p(\theta_j, \theta_{j'})$, $j \neq j'$, while inflating the variances of the elements of $\boldsymbol{\theta}$ on the domain where the variance under q_0 , $\text{Var}_{q_0}(\theta_j)$, must exist for each $j = 1, \dots, d$. The subscripts p and q_0 indicate that the moments are obtained under $p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$ and $q_0(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0)$, respectively. Given a sample \mathbf{Y}_m , possibly a predictor $\mathbf{X}_m = (X_1, \dots, X_m)$, and an ε -information prior $q_0(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0)$, the posterior is

$$q_m(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0, \mathbf{Y}_m, \mathbf{X}_m) \propto q_0(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0) f_m(\mathbf{Y}_m | \mathbf{X}_m, \boldsymbol{\theta}).$$

The ESS is the interpolated value of m minimizing the prior-to-posterior distance δ between $q_m(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0, \mathbf{Y}_m)$ and $p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$.

MTM define the prior-to-posterior distance as the difference between the traces of the information matrix of $p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$ and the expected information matrix of $q_m(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0, \mathbf{Y}_m, \mathbf{X}_m)$, where the expectation is with respect to the prior predictive distribution $f_m(\mathbf{Y}_m | \tilde{\boldsymbol{\theta}}, \boldsymbol{\xi})$. Here, $\bar{\boldsymbol{\theta}} = E_p(\boldsymbol{\theta})$ denotes the prior mean. To compute the distance between $p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$ and $q_m(\boldsymbol{\theta} | \mathbf{Y}_m)$, MTM define

$$\delta(m, \bar{\boldsymbol{\theta}}, p, q_0) = \left| \sum_{j=1}^d D_{p,j}(\bar{\boldsymbol{\theta}}) - \sum_{j=1}^d \int D_{q,j}(m, \bar{\boldsymbol{\theta}}, \mathbf{Y}_m) df_m(\mathbf{Y}_m | \tilde{\boldsymbol{\theta}}) \right|, \tag{2}$$

where $D_{p,j}$ and $D_{q,j}$ are the curvatures of the original prior and the posterior under the ε -information prior, $D_{p,j}(\boldsymbol{\theta}) = -\partial^2 / \partial \theta_j^2 \log p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}})$ and $D_{q,j}(m, \boldsymbol{\theta}, \mathbf{Y}_m) = -\partial^2 / \partial \theta_j^2 \log q_m(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}_0, \mathbf{Y}_m)$.

If interest is focused on a subvector $\boldsymbol{\theta}_r$ of $\boldsymbol{\theta}$, the ESS can be determined similarly in terms of the marginal prior $p(\boldsymbol{\theta}_r | \tilde{\boldsymbol{\theta}})$. When the expectation cannot be obtained analytically, a simulation-based numerical approximation is used. For regression models of \mathbf{Y}_m as a function of a predictor \mathbf{X}_m , the likelihood is $f_m(\mathbf{Y}_m | \boldsymbol{\theta}, \mathbf{X}_m) = \prod_{i=1}^m f(Y_i | \boldsymbol{\theta}, X_i)$. In such settings, MTM augment the model by assuming a sampling model $g_m(\mathbf{X}_m | \boldsymbol{\xi})$ and prior $r(\boldsymbol{\xi} | \tilde{\boldsymbol{\xi}})$, and define

$$f_m(\mathbf{Y}_m | \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\xi}}) = \int f_m(\mathbf{Y}_m | \mathbf{X}_m, \boldsymbol{\theta}) g_m(\mathbf{X}_m | \boldsymbol{\xi}) p(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}) r(\boldsymbol{\xi} | \tilde{\boldsymbol{\xi}}) d\boldsymbol{\theta} d\boldsymbol{\xi}.$$

4 Case Studies of Data Analysis and Study Design

The following examples show how prior sensitivity may be evaluated using ESS in data analysis and clinical trial design settings. The first three, Examples 1 to 3, are data analyses and the latter three, Examples 4 to 6, are clinical trial designs. In each example, we explain how the prior ESS can be used as a tool to calibrate prior hyperparameters. Following Gelman et al. [6], we write $\text{Unif}(\alpha, \beta)$, $\text{Be}(\alpha, \beta)$, $\text{Bin}(n, \theta)$, $\text{Ga}(\alpha, \beta)$, $\text{IG}(\alpha, \beta)$, $\text{Exp}(\theta)$, $\text{N}(\mu, \sigma^2)$, and $\text{MVN}(\mu, \Sigma)$, for the uniform, beta, binomial, gamma, exponential, normal, and multivariate normal distributions.

4.1 Cross-tabulated Counts from Small Samples—Example 1

This example was described earlier, in Sect. 2. The goal of Carlin's analysis was to examine the effects of mechanical ventilator devices on lung damage in rabbits. Because of the very small numbers of studied animals, as shown in Table 1, the effect of a range of non-informative priors was explored. Recall that $\pi_{g,h}$ was the probability of lung damage in cell (g, h) for amplitude g ($= 1, 2$) and frequency h ($= 1, 2, 3$), and $\pi_{g,h}(\boldsymbol{\theta}) = \text{logit}^{-1}(\mu + \alpha_g + \beta_h + \gamma_h I_{(g=2)})$, where the model parameter $\boldsymbol{\theta} = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3)$.

Assuming a binomial model, $Y_{g,h} | \boldsymbol{\theta} \sim \text{Bin}(n_{g,h}, \pi_{g,h}(\boldsymbol{\theta}))$, the likelihood is

$$f(\mathbf{Y}_m | \boldsymbol{\theta}) \propto \prod_{g=1}^2 \prod_{h=1}^3 \pi_{g,h}(\boldsymbol{\theta})^{Y_{g,h}} \{1 - \pi_{g,h}(\boldsymbol{\theta})\}^{n_{g,h} - Y_{g,h}}.$$

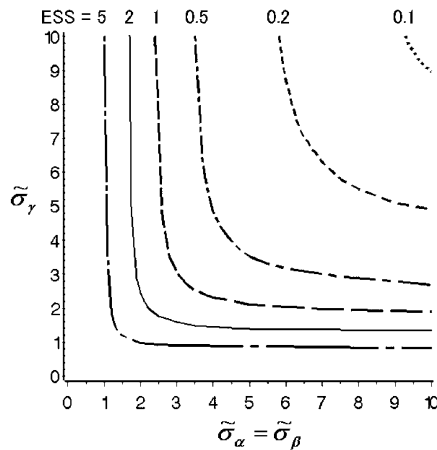
Using MTM's method, we compute an overall ESS, and also ESS_μ , ESS_α , ESS_β , and ESS_γ of the subvectors, which characterize the overall mean, the main effects for amplitude and frequency, and their interaction effects, respectively. The ESS values obtained for the ten alternative choices for the independent normal priors are summarized in Table 2. As discussed in Sect. 2, since the sample size is 38, the prior N10 having the $\text{ESS} = 36.6$ may be criticized as being excessively informative. The priors N7 and N10 both assume very high prior information for the interaction effects. Both priors have $\text{ESS}_\gamma = 96.1$. In this example, the ESS_γ computation shows that the value $\tilde{\sigma}_\gamma = 0.5$ is far too small.

While each of the hyperparameters $\tilde{\sigma}_\alpha$, $\tilde{\sigma}_\beta$, $\tilde{\sigma}_\gamma$ can have an important impact on posterior inferences, they may be difficult to elicit or calibrate. We demonstrate here how one can use prior ESS graphically to assist in choosing these parameters. The idea is to compute the ESS for different combinations of the hyperparameters $\tilde{\boldsymbol{\theta}}$, similarly to Table 2. The ESS values are then plotted as a function of $\tilde{\boldsymbol{\theta}}$. A practical problem is that one must first reduce the dimension of $\tilde{\boldsymbol{\theta}}$ to $d = 2$ to allow one to construct a contour plot of ESS as a function of two hyperparameter values. As a general strategy, we suggest the use of simple restrictions on elements of $\tilde{\boldsymbol{\theta}}$. In this example, we will use $\tilde{\sigma}_\alpha = \tilde{\sigma}_\beta \equiv \tilde{\sigma}$, allowing us to plot ESS contours as a function of the two remaining hyperparameters $(\tilde{\sigma}, \tilde{\sigma}_\gamma)$. Figure 1 shows the contours for $\text{ESS} = 0.1, 0.2, 0.5, 1.0, 2.0$ and 5.0 . From Fig. 1 one might, for example, decide to set $\tilde{\sigma}_\alpha = \tilde{\sigma}_\beta = 2.6$ and $\tilde{\sigma}_\gamma = 4.7$ to obtain $\text{ESS} = 1.0$.

4.2 Bivariate Normal Model for Multiple Correlated Diagnostic Tests—Example 2

Choi et al. [3] used a bivariate normal model to analyze multiple correlated diagnostic tests. They considered the problem of comparing two serologic tests, both enzyme-linked immunosorbent assays (ELISA) for detection of antibodies to Johne's disease in dairy cattle. Data from $n_1 = 88$ diseased animals and $n_0 = 393$ disease-free animals were reported. The two tests have continuous outcomes, which we denote by Y_{1iD} and Y_{2iD} for the i th diseased animal, and by $Y_{1i'\bar{D}}$ and $Y_{2i'\bar{D}}$ for the i 'th disease-free animal. Provided that the same prior and likelihood pair are used for the two independent data sets, they will have the same ESS value. Therefore, hereafter we will

Fig. 1 Contour plots of ESS values for $\tilde{\sigma}_\alpha = \tilde{\sigma}_\beta$ (x-axis) and $\tilde{\sigma}_\gamma$ (y-axis), obtained in Example 1 ([1]). Dotted, dashed, dashed-dotted, longer dashed, solid, longer dashed-dotted lines show the plots of ESS = 0.1, 0.2, 0.5, 1.0, 2.0, and 5.0, respectively



consider the sampling model and prior distributions for one of the two groups, and we drop the subscripts for disease status.

A bivariate normal distribution is assumed for the test scores, (Y_{1i}, Y_{2i}) , from the i th animal. Let $\boldsymbol{\mu} = (\mu_1, \mu_2)$ denote the means and let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\tau_1, \tau_2, \rho)$ denote a 2×2 covariance matrix with marginal variances $(\tau_1^{-1}, \tau_2^{-1})$ and correlation ρ . Choi et al. [3] assume

$$(Y_{i1}, Y_{i2}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

so that $\boldsymbol{\theta} = (\mu_1, \mu_2, \tau_1, \tau_2, \rho)$ and $d = 5$. To complete the model, Choi et al. [3] assume independent prior distributions

$$\mu_j \sim N(\tilde{\mu}_j, \tilde{\tau}_j^{-1}) \quad \text{and} \quad \tau_j \sim \text{Ga}(\tilde{a}_j, \tilde{b}_j), \quad j = 1, 2, \tag{4}$$

with $\tilde{\mu}_j = 0, \tilde{\tau}_j = 0.001, \tilde{a}_j = \tilde{b}_j = 0.001, j = 1, 2,$ and $\rho \sim \text{Unif}(-1, 1)$. The intention is to formalize vague prior information. In anticipation of the upcoming discussion, we use instead a more general scaled Beta prior, $\rho \sim \text{rBe}(\tilde{\alpha}, \tilde{\beta})$. The scaled Beta model $x \sim \text{rBe}(a, b)$ for $L < x < U$ is defined as $(x - L)/(U - L) \sim \text{Be}(a, b)$, i.e., $\rho \sim \text{Unif}(-1, 1)$ for $\tilde{\alpha} = \tilde{\beta} = 1$. We apply MTM's method to compute an overall ESS for $\boldsymbol{\theta} = (\mu_1, \mu_2, \tau_1, \tau_2, \rho)$, and the two additional values, ESS_μ and ESS_Σ , for the subvectors $\boldsymbol{\theta}_\mu = (\mu_1, \mu_2)$ and $\boldsymbol{\theta}_\Sigma = (\tau_1, \tau_2, \rho)$. The computation yields $\text{ESS} < 0.001, \text{ESS}_\mu = 0.001,$ and $\text{ESS}_\Sigma < 0.001$. We interpret these ESSs as evidence of very vague priors, as intended.

To show how ESS may be applied as a tool for prior elicitation in this setting, we consider the four alternative priors shown in Table 3. This serves as an informal sensitivity analysis. Also, similarly to the discussion in Sect. 4.1, we plot ESS as a function of the hyperparameters. We compute the ESS for $\tilde{\tau}_1 = \tilde{\tau}_2$ and $\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2$ each ranging from 0.001 to 10, keeping $\tilde{\alpha} = \tilde{\beta} = 1$ fixed. Figure 2 gives plots of the resulting ESS values. For example, the prior ESS for $\tilde{\tau}_1 = \tilde{\tau}_2 = 10, \tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2 = 10, \tilde{\alpha} = \tilde{\beta} = 1$ is 9.5. This may be criticized as unacceptably high, considering the sample size of $n = 88$ diseased animals. In contrast, priors with all hyperparameters less than 1 correspond to reasonably small prior ESS.

Table 3 Computed ESSs for the correlated diagnostic tests in Choi et al. [3]

$\tilde{\tau}_1 = \tilde{\tau}_2$	$\tilde{a}_1 = \tilde{b}_1 = \tilde{a}_2 = \tilde{b}_2$	$\tilde{\alpha}, \tilde{\beta}$	ESS	ESS $_{\mu}$	ESS $_{\Sigma}$
0.001	0.001	1, 1	<0.001	0.001	<0.001
1	1	1, 1	0.25	1.0	0.14
10	10	1, 1	9.5	10.0	9.1
1	1	5, 5	0.54	1.0	0.53
10	10	5, 5	11.3	10.0	12.4

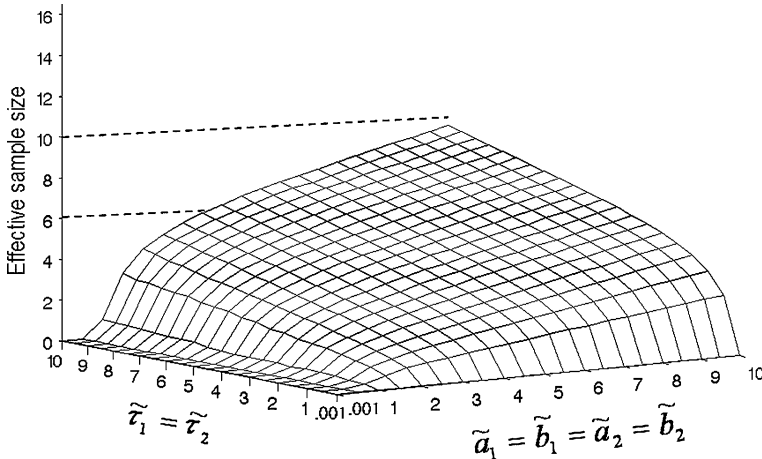


Fig. 2 ESS surface computed for $\tilde{\tau}_1 = \tilde{\tau}_2$ and $\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2$ ranging from 0.001 to 10, keeping $\tilde{\alpha} = \tilde{\beta} = 1$, in Example 2 (Choi et al. [3])

4.3 Proportional-Odds Model for Ordinal Outcomes—Example 3

Congdon [4] (Sect. 10.3.2) reports a data analysis based on a proportional-odds model for ordinal response data, as shown in Table 4. The data report deformity or mortality in chick embryos as a result of arbovirus injection. Two virus groups, Facey’s Paddock ($g = 1$) and Tinaroo ($g = 2$), and a control group ($g = 0$) were investigated. The control group received no virus. The two virus groups and the control group contained $n_1 = 75$, $n_2 = 72$, and $n_0 = 18$ embryos, respectively. Each embryo in the Facey’s Paddock group received one of the four doses, $\{3, 18, 30, 90\}$, denoted by $\{d_{1,1}, d_{1,2}, d_{1,3}, d_{1,4}\}$. For the Tinaroo group, the doses were $\{3, 20, 2400, 88000\}$, denoted by $\{d_{2,1}, d_{2,2}, d_{2,3}, d_{2,4}\}$. The response $Y_{g,i}$ for embryo i in group g was ordinal with three possible values: survival without deformity ($Y = 0$), survival with deformity ($Y = 1$), and death ($Y = 2$).

In this example, a nonzero probability of death was assumed for zero dose. This accounts for a possible background mortality effect. In fact, one death was observed among the controls. The response in the control group $Y_{0,i}$ is assumed to be binary (0 or 2) rather than trinary, with $\Pr(Y_{0,i} = 2 | \alpha) = \alpha$, and $\Pr(Y_{g,i} = h)$ is assumed to be a mixture

$$\pi_{g,i,h} = \Pr(Y_{g,i} = h) = \alpha + (1 - \alpha)P_{g,i,h}, \tag{5}$$

Table 4 Arbovirus injection data (three outcomes: survival without deformity, survival with deformity, and death) reported in Congdon [4]

Virus group	Dose level	Survival without deformity	Survival with deformity	Death	Total
Control	0	17	0	1	18
Facey's Paddock	3	13	1	3	17
	18	14	1	4	19
	30	9	2	8	19
	90	2	1	17	20
Tinaroo	3	18	0	1	19
	20	17	0	2	19
	2400	2	9	4	15
	88000	0	10	9	19

with $P_{g,i,2} = \gamma_{g,i,2}$, $P_{g,i,1} = \gamma_{g,i,1} - \gamma_{g,i,2}$, $P_{g,i,0} = 1 - \gamma_{g,i,1}$, and

$$\gamma_{g,i,h} = \text{logit}^{-1}(\kappa_{g,i} + \beta_g X_{g,i}) \tag{6}$$

for $h = 1, 2$ with $\gamma_{g,i,0} \equiv 1.0$. The covariates are log doses, $X_{g,(z)} = \log_{10}\{d_{g,(z)}\}$, for $z = 1, 2, 3, 4$. In this example, $\theta = (\alpha, \beta_1, \beta_2, \kappa_{1,1}, \kappa_{1,2}, \kappa_{2,1}, \kappa_{2,2})$ and $d = 7$. We consider the two subvectors, $\theta_1 = (\alpha)$ and $\theta_2 = (\beta_1, \beta_2, \kappa_{1,1}, \kappa_{1,2}, \kappa_{2,1}, \kappa_{2,2})$, which characterize the background mortality effect and the dose–response model, respectively. Using dummy indicators $Z_{g,i,h} = 1$ if $Y_{gi} = h$ and 0 otherwise, the likelihood for m_0, m_1 and m_2 embryos in the control group, group 1 and group 2 is

$$f(\mathbf{Z}_m | \theta) \propto \prod_{i_0=1}^{m_0} (1 - \alpha)^{Z_{0,i_0,0}} \alpha^{Z_{0,i_0,2}} \prod_{i_1=1}^{m_1} \pi_{1,i_1,0}^{Z_{1,i_1,0}} \pi_{1,i_1,1}^{Z_{1,i_1,1}} \pi_{1,i_1,2}^{Z_{1,i_1,2}} \prod_{i_2=1}^{m_2} \pi_{2,i_2,0}^{Z_{2,i_2,0}} \pi_{2,i_2,1}^{Z_{2,i_2,1}} \pi_{2,i_2,2}^{Z_{2,i_2,2}}.$$

Congdon [4] assumes independent prior distributions:

$$\alpha \sim \text{Be}(\tilde{\phi}, \tilde{\phi}), \quad \beta_j \sim N(\tilde{\mu}_\beta, \tilde{\sigma}_\beta^2), \quad \text{and} \quad \kappa_{j,h} \sim N(\tilde{\mu}_\kappa, \tilde{\sigma}_\kappa^2),$$

$$j = 1, 2, \quad h = 1, 2,$$

with $\tilde{\phi} = 1$, $\tilde{\mu}_\beta = 0$ and $\tilde{\sigma}_\beta^2 = 10$, $\tilde{\mu}_\kappa = 0$ and $\tilde{\sigma}_\kappa^2 = 100$.

We evaluate the overall ESS for θ , and subvector-specific ESS values ESS_{BG} for θ_1 , and ESS_{DR} for θ_2 . The computations yield $\text{ESS} = 3.3$, $\text{ESS}_{BG} = 3.6$, and $\text{ESS}_{DR} = 0.85$. Compared to the sample size of $n = 165$, the prior distributions used in this example appear appropriately non-informative. Figure 3 shows an ESS_{DR} surface computed for $\tilde{\sigma}_\beta^2$ and $\tilde{\sigma}_\kappa^2$ ranging from 5 to 100, fixing $\tilde{\phi} = 1$. The ESS surface suggests that $\tilde{\sigma}_\beta^2$ and $\tilde{\sigma}_\kappa^2$ both over 50 provide a dose–response model with sufficiently vague priors.

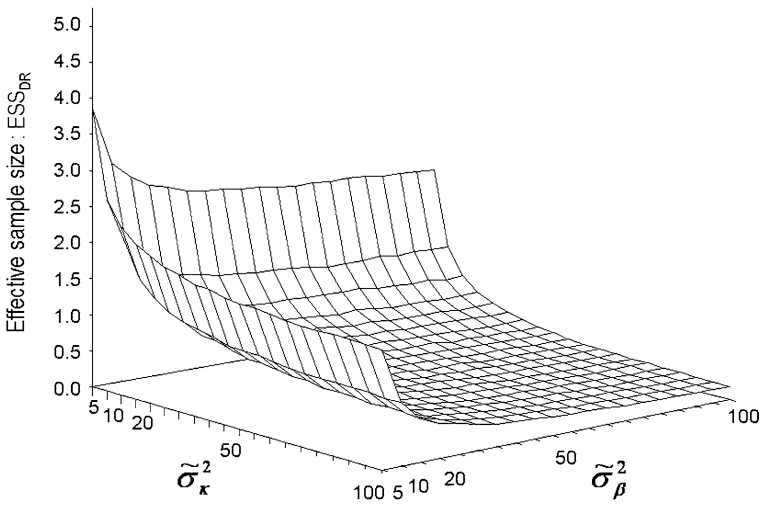


Fig. 3 ESS surface obtained for the priors of the parameters modeling the dose–response relationships, for $\tilde{\sigma}_\beta^2 = \tilde{\sigma}_\kappa^2$ ranging from 5 to 100, fixing $\tilde{\phi} = 1$, in Example 3 (Congdon [4])

4.4 Time-to-Event Continual Reassessment Method—Example 4

The continual reassessment method (CRM) [9] is used for dose-finding in phase I clinical trials based on a binary indicator of toxicity. The CRM requires complete follow-up of the current patient (or cohort) before enrolling a new patient or cohort. Depending on how long it takes to evaluate toxicity, this may lead to an unduly long study duration that make the method impractical. Cheung and Chappell [2] proposed an extension, the time-to-event (TITE) CRM, that uses time to toxicity or right censoring as the outcome.

Elkind et al. [5] applied the TITE-CRM to determine the maximum tolerated dose (MTD) of short-term high-dose lovastatin in stroke patients treated within 24 hours of symptom onset. Each patient received one of five initial doses 1, 3, 6, 8, 10 mg/kg, on days 1–3 post onset and received 20 mg/day for the next 27 days. Toxicity was assessed up to day 30, that is, the observation window was $T_{up} = 30$ days. Denote the time-to-toxicity in patient i by u_i , and the toxicity indicator $Y_i = 1$ if $u_i \leq T_{up}$, 0 if not, so that u_i is right-censored at T_{up} . The dose–toxicity model

$$\Pr(Y_i = 1 \mid d_{[i]}, \beta) = F(d_{[i]}, \beta) = d_{[i]}^{\exp(\beta)}$$

was assumed, where $d_{[i]}$ is the standardized dose level assigned to patient i . An $N(0, 1.34)$ distribution was assumed for the prior of β . The five standardized doses $\mathbf{d} = (d_1, d_2, d_3, d_4, d_5)$ in the model were assumed to be (0.02, 0.06, 0.10, 0.18, 0.30). In general, the TITE-CRM is implemented using the weighted working likelihood for m patients given by

$$f_m(\mathbf{Y}_m, \mathbf{u}_m \mid \mathbf{d}_m, \beta) = \prod_{i=1}^m F(d_{[i]}, \beta)^{Y_i} \{1 - w_i F(d_{[i]}, \beta)\}^{1-Y_i}, \tag{7}$$

Table 5 Toxicity scenarios for a dose–toxicity relationship (TITE-CRM example)

Scenario	1 mg/kg	3 mg/kg	6 mg/kg	8 mg/kg	10 mg/kg
(1)	0.02	0.06	0.10	0.18	0.30
(2)	0.10	0.20	0.35	0.55	0.70
(3)	0.30	0.35	0.50	0.70	0.80
(4)	0.10	0.50	0.80	0.90	0.90
(5)	0.40	0.70	0.85	0.90	0.90

where w_i is a suitable weight function. For the lovastatin trial, $w_i = u_i / T_{up}$ was used, and patients were assumed to arrive according to a Poisson process. This is equivalent to assuming that the inter-arrival times are i.i.d. $\text{Exp}(\lambda)$. In the trial, $\lambda = 2$ patients per month was assumed.

In Table 5, we assume five dose–toxicity scenarios in order to assess effects of the prior ESS. Scenario (1) corresponds to toxicity probabilities equal to the standardized doses. Scenarios (2)–(5) were constructed by starting with Scenario (1) and increasing the toxicity probabilities. In Scenarios (2) and (3), the toxicity probabilities increase linearly with dose, with all doses too toxic in Scenario (3). In Scenario (4), only d_1 is safe, with toxicity increasing rapidly from d_2 onward. In Scenario (5), all doses are very toxic.

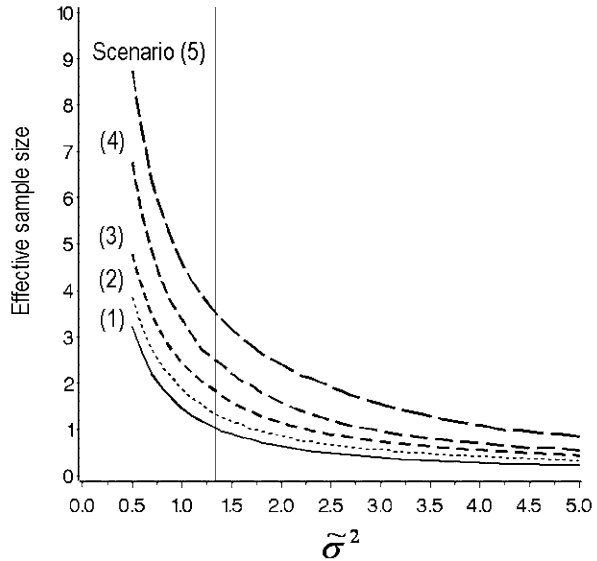
As Cheung and Chappell [2] do, we assume three models for the patients' times to toxicity, including a conditionally uniform model, a Weibull model (with a fixed shape parameter 4), and a log-logistic model (with a fixed shape parameter 1). The cumulative distribution function (CDF) of the Weibull model with a scale parameter α is $F(u, \alpha) = 1 - \exp\{-(u/\alpha)^4\}$ and the CDF of the log-logistic model with a scale parameter α is $F(u, \alpha) = (1 + \exp[-\{\log(u) - \log(\alpha)\}])^{-1}$.

We compute the ESS values under each model. Figure 4 gives plots of the ESS values as a function of $\tilde{\sigma}^2$ under the five toxicity scenarios, assuming the conditionally uniform model for time-to-toxicity and with the prior mean of β fixed at $\tilde{\mu} = 0$. Since the ESS computed at $\tilde{\sigma}^2 = 1.34$ is less than 2 under Scenario (1), the information from the likelihood will dominate the prior after enrolling 3 patients, hence the prior specified in the lovastatin trial seems quite reasonable. The prior also makes sense under Scenarios (2) and (3). The plot of ESS under Scenarios (4) and (5) indicates that the prior may be problematic, however. Under Scenario (5), it appears that $\tilde{\sigma}^2 > 2.5$ may be needed to ensure an ESS < 2. The findings are similar under the Weibull and log-logistic models. This example illustrates that prior ESS computations can be a useful device to help calibrate the prior to improve the behavior of the TITE-CRM.

4.5 Trial Monitoring for Time-to-Event Outcomes—Example 5

Thall et al. [11] present a series of study designs for monitoring time-to-event outcomes in early phase clinical trials. We focus on one of the study designs, which was applied to a single-arm phase II trial for advanced kidney cancer. In the trial, the plan was to enroll up to 84 patients, with each patient's disease status evaluated up to 12 months. In this example, we focus on the mean time-to-event, μ . For patient i , let T_i

Fig. 4 Plots of ESS values against $\tilde{\sigma}^2$ under the five toxicity scenarios given in Table 5, in Example 4 (Cheung and Chappell [2]). The vertical line at $\tilde{\sigma}^2 = 1.34$ indicates the hyperparameter value that was actually used in the lovastatin trial



denote the time to disease progression (failure), let T_i^o be the observed value of T_i or the administrative right-censoring time, and let $Y_i = I(T_i^o = T_i)$. We assume the t_i 's are i.i.d. $\text{Exp}(\mu)$, exponential with mean μ , which has pdf $f(t | \mu) = \mu^{-1} \exp(-t/\mu)$ and survivor function $F(t | \mu) = \Pr(T > t | \mu) = \exp(-t/\mu)$. The likelihood for m patients is

$$f_m(\mathbf{Y}_m, \mathbf{T}_m^o | \mu) = \mu^{-\sum_{i=1}^m Y_i} \exp\left(-\sum_{i=1}^m T_i^o / \mu\right). \tag{8}$$

Using the relationship $\mu = \text{mean}(T) = \text{median}(T) / \log(2)$, Thall et al. [11] established the prior of μ_S corresponding to the historical standard treatment from elicited mean values and a 95% credible interval of $\text{median}(T)$. This gave an inverse gamma $\text{IG}(\tilde{\alpha}, \tilde{\beta})$ with $(\tilde{\alpha}, \tilde{\beta}) = (53.477, 301.61)$ as the prior $p(\mu_S | \tilde{\mu}_S)$. Here, $\text{IG}(\tilde{\alpha}, \tilde{\beta})$ denotes an inverse gamma distribution with mean $\tilde{\beta} / (\tilde{\alpha} - 1)$ and variance $\tilde{\beta}^2 / \{(\tilde{\alpha} - 1)^2(\tilde{\alpha} - 2)\}$, which requires $\tilde{\alpha} > 2$. The prior of μ_E in the experimental treatment $p(\mu_E | \tilde{\mu}_E)$ was calibrated to have the same mean but inflated variance to reflect much greater prior uncertainty about the experimental treatment. This yielded $\mu_E \sim \text{IG}(5.348, 30.161)$. The time scale of the event time and its corresponding parameter μ are in months.

The prior ESS in a simple inverse gamma-exponential model with an inverse gamma prior, $\mu \sim \text{IG}(\tilde{\alpha}, \tilde{\beta})$, and the exponential sampling model, $T \sim \text{Exp}(\mu)$, is analytically determined to be $\tilde{\alpha} - 2$. Thus, the ESS of the $\text{IG}(5.348, 30.161)$ prior is 3.348 under this model. This prior ESS is obtained under the assumption that T is observed for all accrued patients; that is, no censoring occurs. Since in general the ESS is defined as a property of a prior and likelihood pair, a given prior might have different ESS values for different likelihoods. As mentioned in Sect. 3, our approach defines the ESS to be the sample size that yields a posterior containing the same

amount of information as the prior. It is well known that the amount of information for time-to-event data depends on the number of observed events, not the sample size. Therefore, when T_i is right-censored for some patients, the prior ESS should be larger than $\tilde{\alpha} - 2$. We apply MTM's method to compute an ESS under the inverse gamma prior $\mu_E \sim \text{IG}(5.348, 30.161)$ with respect to the likelihood (8). The computation yields $\text{ESS} = 4.8$.

In clinical trials with Bayesian adaptive decision making, it is important to evaluate the impact of the prior on the stopping rule. In the study design of Thall et al. [11], the trial should be stopped early if, based on the current data,

$$\Pr(\mu_S + 4.3 < \mu_E \mid \text{data}) < 0.015. \tag{9}$$

This rule stops the trial if it is unlikely that the mean failure time with the experimental treatment is at least a 4.3 month improvement over the historical mean with the standard treatment. The 4.3 month improvement in mean failure time corresponds to a 3.0 month improvement in median failure time, since $4.3 = 3.0/\log(2)$. In order to evaluate the impact of the prior $p(\mu_E \mid \tilde{\mu}_E)$ on the stopping rule, we simulated the trial under each of a set of priors having different variances, corresponding to prior ESS values ranging from 1 to 20. We generated exponential patient event times using fixed (true) parameters $\mu_E^{\text{true}} = 5.7, 7.2, 8.6, 10.0$, which correspond to median failure times of 4, 5, 6, 7 months. Each case was simulated 2000 times.

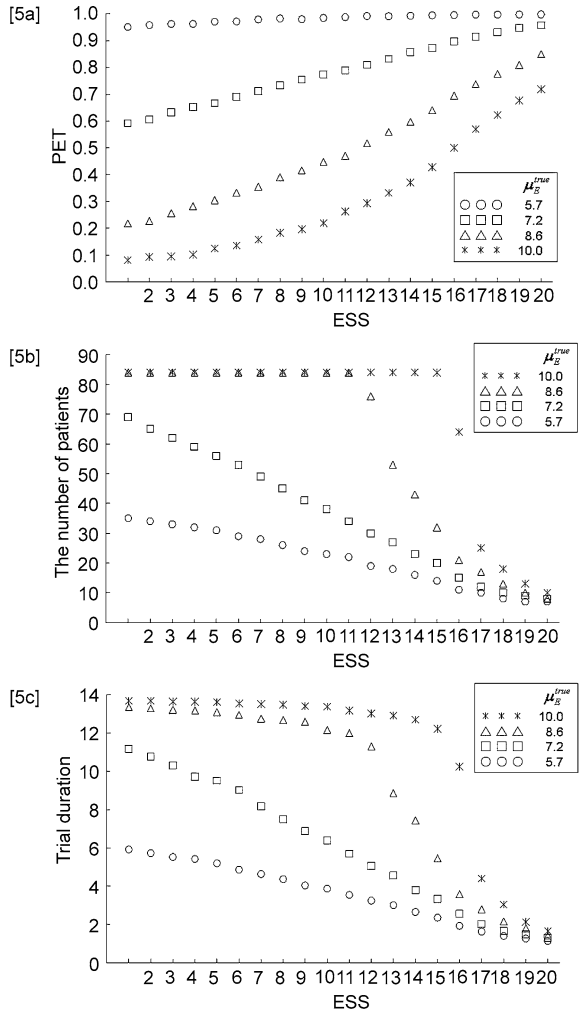
Figures 5a, 5b, and 5c illustrate the simulation results in terms of the probability of early termination (PET), the number of patients, and trial duration, respectively. Figure 5a shows plots of PET as a function of ESS for four values of μ_E^{true} . Since the prior mean of μ_S under $\text{IG}(53.477, 301.61)$ is 5.7, the PET values obtained under the four μ_E^{true} values are reasonable for ESS values up to about 10. In contrast, for $\text{ESS} > 15$, the prior, rather than the data, dominates early stopping decisions. With respect to the number of patients and trial duration, plots of the 50th percentiles of their distributions are shown in Figs. 5b and 5c, respectively. The same findings as with PET are observed; that is, a prior $\text{ESS} > 15$ may be excessively informative.

4.6 A Dose–Response Model for Bivariate Binary Outcomes—Example 6

Thall and Cook [10] use a bivariate binary regression model in a dose-finding trial where each patient is treated at one of four doses $\{0.25, 0.50, 0.75, 1.00\}$ mg/m². Denoting these by d_1, d_2, d_3, d_4 , the standardized doses $X_{(z)} = \log(d_z) - (1/4) \sum_{e=1}^4 \log(d_e)$ are used in the model. Let $Y = (Y_E, Y_T)$ be indicators of efficacy and toxicity, and let $\pi_{a,b}(X, \theta) = \Pr(Y_E = a, Y_T = b \mid X, \theta)$ for $a, b \in \{0, 1\}$. The marginal probabilities are modeled as $\pi_k(X, \theta_k) = \text{logit}^{-1}\{\eta_k(X, \theta_k)\}$ for $k = E, T$, with linear predictors $\eta_E(X, \theta_E) = \mu_E + X\beta_{E,1} + X^2\beta_{E,2}$ and $\eta_T(X, \theta) = \mu_T + X\beta_T$. The joint probabilities $\pi_{a,b}$ are modeled in terms of these marginal probabilities and one real-valued association parameter ψ :

$$\pi_{a,b} = \pi_E^a(1 - \pi_E)^{1-a}\pi_T^b(1 - \pi_T)^{1-b} + (-1)^{a+b}\pi_E(1 - \pi_E)\pi_T(1 - \pi_T)\left(\frac{e^\psi - 1}{e^\psi + 1}\right), \tag{10}$$

Fig. 5 Plots of (a) PET and the 50th percentile points of the distributions of (b) the number of patients and (c) trial duration, under $\mu_E^{true} = 5.7$ (circle), 7.2 (square), 8.6 (triangle), and 10.0 (star) against prior ESS (Example 5: Thall et al. [11])



for $a, b \in \{0, 1\}$. Thus, $\theta = (\mu_E, \beta_{E,1}, \beta_{E,2}, \mu_T, \beta_T, \psi)$ and $d = 6$. The likelihood for m patients is

$$f(\mathbf{Y}_m | \mathbf{X}_m, \theta) = \prod_{i=1}^m \prod_{a=0}^1 \prod_{b=0}^1 \pi_{a,b}(X_i, \theta)^{I\{Y_i=(a,b)\}}.$$

The prior $p(\theta | \tilde{\theta})$ was established from elicited mean values of $\pi_E(X, \theta)$ and $\pi_T(X, \theta)$, which yielded normal distributions with hyperparameters $(\tilde{\mu}_{\mu_E}, \tilde{\sigma}_{\mu_E}^2) = (-1.496, 1.113^2)$, $(\tilde{\mu}_{\beta_{E,1}}, \tilde{\sigma}_{\beta_{E,1}}^2) = (1.180, 0.869^2)$, $(\tilde{\mu}_{\beta_{E,2}}, \tilde{\sigma}_{\beta_{E,2}}^2) = (0.149, 1.192^2)$, $(\tilde{\mu}_{\mu_T}, \tilde{\sigma}_{\mu_T}^2) = (-0.619, 0.941^2)$, $(\tilde{\mu}_{\beta_T}, \tilde{\sigma}_{\beta_T}^2) = (0.587, 1.659^2)$, and $(\tilde{\mu}_{\psi}, \tilde{\sigma}_{\psi}^2) = (0, 10)$ where $\tilde{\sigma}_{\psi}^2$ is modified for this illustration. We apply MTM's method to compute the ESS of $p(\theta | \tilde{\theta})$, and ESS_E , ESS_T , and ESS_{ψ} for the subvectors of the efficacy

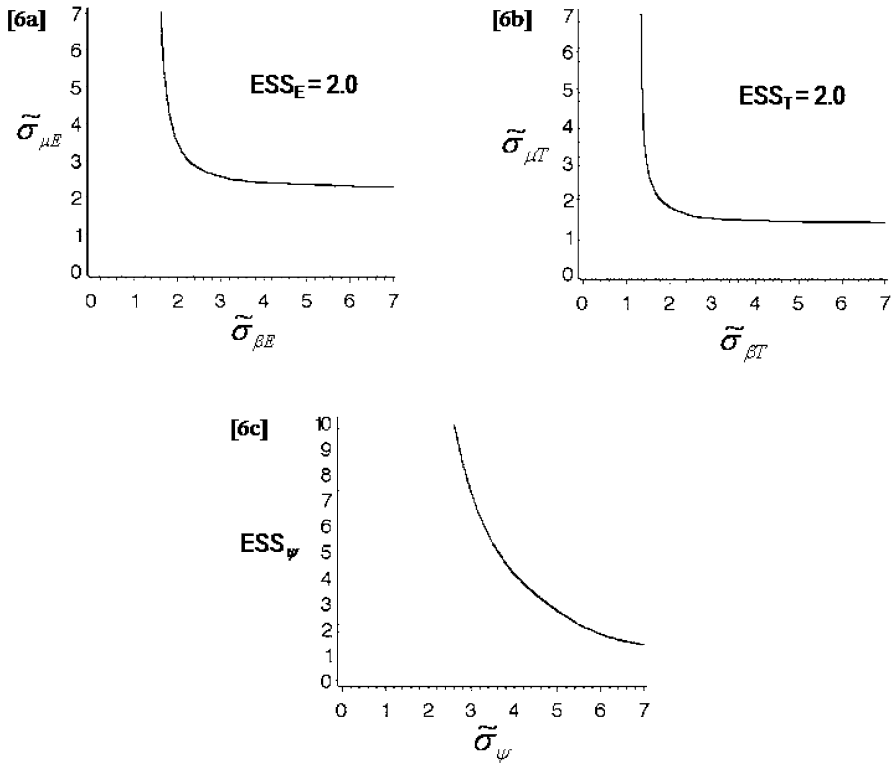


Fig. 6 (a) Contour plot of $ESS_E = 2.0$, fixing $(\tilde{\sigma}_{\mu_T}, \tilde{\sigma}_{\beta_T}, \tilde{\sigma}_\psi) = (0.941, 1.659, 1)$. (b) Contour plot of $ESS_T = 2.0$, fixing $(\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E}, \tilde{\sigma}_\psi) = (2.68, 2.68, 3.16)$. (c) Plot of ESS_ψ values against $\tilde{\sigma}_\psi$, fixing $\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E}, \tilde{\sigma}_{\mu_T}, \tilde{\sigma}_{\beta_T} = (2.68, 2.68, 1.89, 1.89)$. The three plots are obtained in Example 6 (Thall and Cook [10])

parameters $\theta_E = (\mu_E, \beta_{E,1}, \beta_{E,2})$, toxicity parameters $\theta_T = (\mu_T, \beta_T)$, and the association parameter ψ . The computations yield $ESS = 8.9$, $ESS_E = 13.7$, $ESS_T = 5.3$, and $ESS_\psi = 9.0$.

ESS values computed for the subvectors of the parameters, as well as the full parameter vector, are a useful feedback in the prior elicitation process. We assume fixed hyperparameters $\tilde{\mu}_{\mu_E}, \tilde{\mu}_{\beta_{E,1}}, \tilde{\mu}_{\beta_{E,2}}, \tilde{\mu}_{\mu_T}, \tilde{\mu}_{\beta_T}$, and $\tilde{\mu}_\psi$, and discuss the choice of the variance parameters $\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_{E1}}, \tilde{\sigma}_{\beta_{E2}}, \tilde{\sigma}_{\mu_T}, \tilde{\sigma}_{\beta_T}$, and $\tilde{\sigma}_\psi$. We demonstrate how to calibrate the priors in this example. In the design described by Thall and Cook [10], up to $N = 36$ patients are treated in cohorts of size = 3. Therefore, it may be desirable that the overall ESS and the subvector ESSs are at most 2, so that the accumulating data dominates the posterior inferences after enrolling 3 patients.

Figure 6 shows the contours of ESS_E , ESS_T , and ESS_ψ . In order to plot a contour of ESS_E for $\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_{E1}}, \tilde{\sigma}_{\beta_{E2}}$, we constrain $\tilde{\sigma}_{\beta_{E1}}, \tilde{\sigma}_{\beta_{E2}} \equiv \tilde{\sigma}_{\beta_E}$ and fix $\{\tilde{\sigma}_{\mu_T}, \tilde{\sigma}_{\beta_T}, \tilde{\sigma}_\psi\} = \{0.941, 1.659, 3.162\}$. This allows us to plot ESS as a function of the two: $(\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E})$. Figure 6a shows the contour for $ESS_E = 2.0$. Inspection of the ESS contours provides a basis for an informed choice of the hyperparameters $(\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E})$. For example,

$\tilde{\sigma}_{\mu_E} = \tilde{\sigma}_{\beta_{E,1}} = \tilde{\sigma}_{\beta_{E,2}} = 2.68$ may be chosen, to ensure that $ESS_E = 2.0$. Similarly, we plot the contour for $ESS_T = 2.0$, fixing $(\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E}, \tilde{\sigma}_{\psi}) = (2.68, 2.68, 3.16)$, shown in Fig. 6b. Inspecting this curve, one may choose values of $\tilde{\sigma}_{\mu_T}, \tilde{\sigma}_{\beta_T}$, for example, $\tilde{\sigma}_{\mu_T} = \tilde{\sigma}_{\beta_T} = 1.89$. We next focus on the interaction parameter and compute ESS_{ψ} values for a range of $\tilde{\sigma}_{\psi}$. The plot is shown in Fig. 6c. One may choose, for example, $\tilde{\sigma}_{\psi} = 5.70$, which provides $ESS_{\psi} = 2.0$. The overall ESS value for the hyperparameter values chosen above is 2.0. If desired, one may repeat the procedure, starting with the choice of $(\tilde{\sigma}_{\mu_E}, \tilde{\sigma}_{\beta_E})$, until a satisfactory overall ESS is obtained. In a last step one may drop the constraint on $\tilde{\sigma}_{\beta_{E1}} = \tilde{\sigma}_{\beta_{E2}}$ and allow different values for these two parameters.

5 Discussion

We have discussed prior sensitivity analyses in Bayesian biostatistics by using prior ESS, illustrated by examples of data analysis and study design for biomedical studies. The main advantage of using ESS is practical feasibility. The definition is pragmatic and allows one to report a meaningful prior ESS summary for most problems. Another important feature is ease of communication. A user need not understand the mathematical underpinnings of the approach to interpret the final report, since the ESS is a hypothetical sample size, in terms of patients (or animals or experimental units), which is readily interpretable.

The ESS provides a numerical value for the effective sample size of a given prior. If one wishes to utilize this methodology to construct a prior having a given ESS, two important cases may be identified. When designing a small to moderate sized clinical trial using Bayesian methods, it is desirable that the prior ESS be small enough so that early decisions are dominated by the data (e.g. the first cohort of 3 patients in a dose-finding study) rather than the prior. In this case, an ESS in the range 0.5 to 2.0 may be appropriate. On the other hand, if one is eliciting a prior for analysis of a given data set of n observations, then a desirable ESS may be specified relative to n . In this case, an ESS of $0.10 \times n$ or smaller might be appropriate.

Morita, Thall, and Müller (MTM2) [8] develop a variation of the ESS suitable for conditionally independent hierarchical models (CIHMs). For a two-level CIHM with K subgroups, in the first level, \mathbf{Y}_k follows distribution $f(\mathbf{Y}_k | \boldsymbol{\theta}_k)$, the subgroup-specific parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ are i.i.d. with prior $\pi_1(\boldsymbol{\theta}_k | \tilde{\boldsymbol{\theta}})$, and the hyperparameter $\tilde{\boldsymbol{\theta}}$ has a hyperprior $\pi_2(\tilde{\boldsymbol{\theta}} | \boldsymbol{\phi})$ with known $\boldsymbol{\phi}$. MTM2 define ESS under a CIHM in two cases, focusing on either the first level prior or second level prior, in order to address different inferential objectives. In case 1, the target is the marginalized prior, $\pi_{12}(\boldsymbol{\theta} | \boldsymbol{\phi}) = \int \pi_1(\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}) \pi_2(\tilde{\boldsymbol{\theta}} | \boldsymbol{\phi}) d\tilde{\boldsymbol{\theta}}$, which may be of interest, for example, if $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$ are the treatment effects in K different canine breeds in a dietary study. In case 2, the target prior is $\pi_2(\tilde{\boldsymbol{\theta}} | \boldsymbol{\phi})$, which would be the focus if the parameter of primary interest is an overall effect $\tilde{\boldsymbol{\theta}}$ for canines, obtained by averaging over the K breeds.

Some important limitations remain. The methodology is based on comparing curvatures of the marginal prior and the posterior distribution under an ε -information prior. Consequently, when an analytic solution does not exist, a limitation is computational complexity. While the actual computational effort is negligible, the choice

of a suitable ε -information prior and the evaluation of the prior–posterior distance require some problem-specific input from the investigator. That is, it is difficult to completely automate the ESS evaluation. However, the examples given here are intended to provide a basis for interested readers to compute and utilize prior ESS in similar problems. A computer program, `ESS_RegressionCalculator.R`, to calculate the ESS for a normal linear or logistic regression model is available from the website <http://biostatistics.mdanderson.org/SoftwareDownload>.

Acknowledgements Satoshi Morita's work was supported in part by Grant H21-CLINRES-G-009 from the Ministry of Health, Labour, and Welfare in Japan. Peter Thall's work was partially supported by Grant NIH/NCI 2R01 CA083932. Peter Müller's work was partially supported by Grant NIH/NCI R01 CA75981.

References

1. Carlin JB (2002) Assessing the homogeneity of three odds ratios: A case study in small- sample inference. In: Gatsonis C, Robert EK, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M (eds) *Case studies in Bayesian statistics*, vol V. Springer, New York, pp 279–290
2. Cheung YK, Chappell R (2000) Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 56:1177–1182
3. Choi YK, Johnson WO, Collins MT, Gardner IA (2006) Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *J Agric Biol Environ Stat* 11:210–229
4. Congdon P (2003) *Applied Bayesian modelling*. Wiley, Chichester
5. Elkind MS, Sacco RL, MacArthur RB, Fink DJ, Peerschke E, Andrews H, Neils G, Stillman J, Corporan T, Leifer D, Cheung K (2008) The neuroprotection with statin therapy for acute recovery trial (NeuSTART): An adaptive design phase I dose-escalation study of high-dose lovastatin in acute ischemic stroke. *Int J Stroke* 3:210–218
6. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian data analysis*, 2nd edn. Chapman and Hall/CRC, New York
7. Morita S, Thall PF, Müller P (2008) Determining the effective sample size of a parametric prior. *Biometrics* 64:595–602
8. Morita S, Thall PF, Müller P (2009) Prior effective sample size in conditionally independent hierarchical models. Technical Report, Yokohama City University
9. O'Quigley J, Pepe M, Fisher L (1990) Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics* 46:33–48
10. Thall PF, Cook JD (2004) Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 60:684–693
11. Thall PF, Wooten LH, Tannir NM (2005) Monitoring event times in early phase clinical trials: Some practical issues. *Clin Trials* 2:467–478