

Supplementary Materials for Prior Effective Sample Size in Conditionally Independent Hierarchical Models

by Satoshi Morita, Peter F. Thall and Peter Müller

Evaluation of the ESS for Special CIHMs

The following provides computational details for the evaluation of the ESS for each of the three examples of widely used CIHMs. The first CIHM is given briefly in Section 5. We write $\text{Be}(\alpha, \beta)$, $\text{Bin}(n, \theta)$, $\text{Ga}(\alpha, \beta)$, $\text{IG}(\alpha, \beta)$, $\text{Exp}(\theta)$, $\text{N}(\mu, \sigma^2)$, $\text{MVN}(\mu, \Sigma)$, and $\text{Inv-}\chi^2(\nu, s)$ for the beta, binomial, gamma, inverse gamma, exponential, normal, multivariate normal, and scaled inverse- χ^2 distributions. In the sequel, we will write “ $\pi_2 - \pi_1 - f$ ”, with specific distributions substituted, to denote a CIHM with hyperprior π_2 , prior π_1 and sampling model f .

1. Normal/Inverse χ^2 -Normal-Normal Model (Section 5)

Case 1: ESS of π_{12} given prior variance $\tilde{\gamma}^2$

Let \mathbf{I} denote the $K \times K$ unit identity and \mathbf{J} the $K \times K$ matrix with all ones. The target prior is specified analytically as

$$\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \tilde{\gamma}^2 \sim \text{MVN}(\boldsymbol{\mu}_{\phi}, \tilde{\gamma}^2 \mathbf{I} + \tau_{\phi}^2 \mathbf{J}),$$

where $\boldsymbol{\mu}_{\phi}$ denotes a $K \times 1$ vector with all entries μ_{ϕ} . The ε -information prior is as given in Section 3.2. First, $\pi_{1,0}(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}, c)$ is specified as $\theta_k \sim \text{i.i.d. } N(\tilde{\mu}, c\tilde{\gamma}^2)$ so that $\pi_{1,0}$ has the same mean as the level 1 prior π_1 but the inflated variance. Then, $\pi_{12,0}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \tilde{\gamma}^2, c)$ is specified as $\boldsymbol{\theta} \sim \text{MVN}(\boldsymbol{\mu}_{\phi}, c\tilde{\gamma}^2 \mathbf{I} + \tau_{\phi}^2 \mathbf{J})$, by marginalizing $\pi_{1,0}$ with respect to the level 2 hyperprior $\tilde{\mu} \mid \mu_{\phi}, \tau_{\phi}^2 \sim N(\mu_{\phi}, \tau_{\phi}^2)$. Note that $\text{Corr}_{\pi_{12,0}}(\theta_k, \theta_{k'})$

$= \tau_\phi^2 / (c\tilde{\gamma}^2 + \tau_\phi^2) \neq \text{Corr}_{\pi_{12}}(\theta_k, \theta_{k'}) = \tau_\phi^2 / (\tilde{\gamma}^2 + \tau_\phi^2)$. In this setting, $-H_{\pi_{12}}(\bar{\boldsymbol{\theta}}_{12}) = (\tilde{\gamma}^2 \mathbf{I} + \tau_\phi^2 \mathbf{J})^{-1}$ and $-H_{\pi_{12,M}}(\bar{\boldsymbol{\theta}}_{12}, \bar{\mathcal{Y}}_M) \approx \left(\frac{\sigma^2}{m} \mathbf{I}\right)^{-1}$, where $m = M/K$ and $\bar{\boldsymbol{\theta}}_{12}$ denotes a vector of the prior mean of θ_k under π_{12} , $\bar{\theta}_{12} = \mu_\phi$. Letting $r = \frac{\tau_\phi^2}{\tilde{\gamma}^2 + \tau_\phi^2}$ denote the intra-class correlation (ICC) under the model assumed here, $-H_{\pi_{12}}(\bar{\boldsymbol{\theta}}_{12}) = (\tilde{\gamma}^2 \mathbf{I} + \tau_\phi^2 \mathbf{J})^{-1} = [(\tilde{\gamma}^2 + \tau_\phi^2) \{(1-r)\mathbf{I} + r\mathbf{J}\}]^{-1}$. The inverse matrix and determinant of $(1-r)\mathbf{I} + r\mathbf{J}$ are respectively given as

$$\frac{1}{1-r} \left\{ \mathbf{I} - \frac{r}{1+(K-1)r} \mathbf{J} \right\} \quad \text{and} \quad \{1 + (K-1)r\}(1-r)^{K-1}.$$

We have approximated $-H_{\pi_{12,M}}(\bar{\boldsymbol{\theta}}_{12}, \bar{\mathcal{Y}}_M)$ by $\left(\frac{\sigma^2}{m} \mathbf{I}\right)^{-1}$, because all elements of $-H_{\pi_{12,0}}$, the minus second derivative matrix of $\log(\pi_{12,0})$, are negligible due to the large positive value c . By computing $\Delta_1(M, \pi_1, \pi_2, \pi_{1,0})$, the ESS is found to be

$$\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}) = \frac{\sigma^2}{\tilde{\gamma}^2} \left(\frac{1-r}{1+(K-1)r} \right)^{1/K} \times K. \quad (\text{S1})$$

In this setting where $\pi_{12,0}$ is specified as a multivariate normal distribution, all elements of $-H_{\pi_{12,0}}$ are negligible. However, this does not always occur in other settings.

Case 1: ESS of π_{12} with random $\tilde{\gamma}^2$.

Allowing $\tilde{\gamma}^2$ to be random, one may obtain a closed form of $\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}^2})$ in the following three steps. First, the distribution of $\pi_1(\boldsymbol{\theta} \mid \tilde{\mu}, \tilde{\gamma}^2) \times \pi_2(\tilde{\gamma}^2 \mid \boldsymbol{\phi}_{\tilde{\gamma}^2})$ marginalized with respect to $\tilde{\gamma}^2$ is a multivariate t distribution with ν_ϕ of degree of freedom and $K \times K$ scale matrix $\mathbf{S} = S_\phi \mathbf{I}$. Second, we approximate this multivariate t distribution with a K -variate normal distribution having mean vector $\tilde{\boldsymbol{\mu}}$ and variance-covariance matrix $\frac{\nu_\phi}{\nu_\phi - 2} \mathbf{S}$, denoted by $\text{MVN}(\tilde{\boldsymbol{\mu}}, \frac{\nu_\phi}{\nu_\phi - 2} \mathbf{S})$. Third, we marginalize this with respect to $\pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}})$, $\tilde{\mu} \mid \mu_\phi, \tau_\phi^2 \sim \text{N}(\mu_\phi, \tau_\phi^2)$, over $\tilde{\mu}$. Denoting the prior mean of $\tilde{\gamma}^2$ under $\pi_2(\tilde{\gamma}^2 \mid \boldsymbol{\phi}_{\tilde{\gamma}^2})$ by $\mu_{\tilde{\gamma}^2} = \frac{\nu_\phi S_\phi}{\nu_\phi - 2}$, this yields the multivariate normal target prior $\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}^2})$,

$$\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}^2} \sim \text{MVN}(\boldsymbol{\mu}_\phi, \mu_{\tilde{\gamma}^2} \mathbf{I} + \tau_\phi^2 \mathbf{J}).$$

The ε -information priors are specified as $\pi_{1,0}(\boldsymbol{\theta} \mid \tilde{\mu}, \tilde{\gamma}^2, c)$, $\boldsymbol{\theta} \mid \tilde{\mu}, \tilde{\gamma}^2, c \sim \prod_{k=1}^K \mathcal{N}(\theta_k \mid \tilde{\mu}, c\tilde{\gamma}^2)$, and $\pi_{12,0}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}^2}, c)$, $\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}^2}, c \sim \text{MVN}(\boldsymbol{\mu}_{\phi}, c\mu\tilde{\gamma}^2\mathbf{I} + \tau_{\phi}^2\mathbf{J})$. As in the previous section, computing $\Delta_1(M, \pi_1, \pi_2, \pi_{1,0})$, gives the ESS

$$\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}) = \frac{\sigma^2}{\mu\tilde{\gamma}^2} \left(\frac{1-r}{1+(K-1)r} \right)^{1/K} \times K, \quad (\text{S2})$$

where r is the ICC equal to $\frac{\tau_{\phi}^2}{\mu\tilde{\gamma}^2 + \tau_{\phi}^2}$.

Case 2: ESS of π_2

We apply Algorithm 2 described in Section 3.3 to $\pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}})$, in order to compute $\Delta_2(M, \pi_2, \pi_{2,0})$, Using the marginal likelihood

$$f_M(\mathcal{Y}_M \mid \tilde{\mu}, \tilde{\gamma}^2, \sigma^2) = \prod_{k=1}^K \mathcal{N}(\bar{Y}_k \mid \tilde{\mu}, \frac{\sigma^2}{m} + \tilde{\gamma}^2)$$

where $\bar{Y}_k = m^{-1} \prod_{i=1}^m Y_{k,i}$, one may compute analytically $-H_{\pi_2}(\bar{\boldsymbol{\theta}}_2) = \tau_{\phi}^{-2}$ and similarly $-H_{\pi_2}(\bar{\boldsymbol{\theta}}_2, \bar{\mathcal{Y}}_{M,2}) \approx K \left(\frac{\sigma^2}{m} + \mu\tilde{\gamma}^2 \right)^{-1}$, where $-H_{\pi_2}(\bar{\boldsymbol{\theta}}_2)$ and $-H_{\pi_2}(\bar{\boldsymbol{\theta}}_2, \bar{\mathcal{Y}}_{M,2})$ denote the minus second derivative of $\log\{\pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}})\}$ and the minus second derivative of $\log\{\pi_{2,M}(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \bar{\mathcal{Y}}_{M,2})\}$ at the prior mean of $\tilde{\boldsymbol{\theta}}$, respectively. Thus,

$$\Delta_2(M, \pi_2, \pi_{2,0}) = \left| \tau_{\phi}^{-2} - K \left(\frac{\sigma^2}{m} + \mu\tilde{\gamma}^2 \right)^{-1} \right|.$$

The ESS is obtained as

$$\text{ESS}_2(\tilde{\mu} \mid \boldsymbol{\phi}) = \frac{\sigma^2}{\tau_{\phi}^2} \left(1 - \frac{\mu\tilde{\gamma}^2}{\tau_{\phi}^2 K} \right)^{-1},$$

under the restriction that $\tau_{\phi}^2 > K^{-1}\mu\tilde{\gamma}^2$. In the above derivation, $-H_{\pi_2}(\bar{\boldsymbol{\theta}}_2, \bar{\mathcal{Y}}_{M,2})$ has the upper bound $K\mu\tilde{\gamma}^2$, however large m becomes. This means that, given K , one cannot increase the amount of information to learn about the overall effect by increasing the

number of subjects in each subgroup, m . Rather one needs a larger number of subgroups to learn more about the overall effect. The number of subgroups K should be sufficiently large so that the amount of information given by the likelihood can exceed that of the prior by increasing m . Otherwise, if the restriction is violated, π_2 has excess information and should be downweighted.

2. Gamma/Inverse Gamma-Gamma-Exponential Model

We next consider a gamma/inverse gamma-gamma-exponential model. This could be used as an alternative hierarchical model for the multicenter randomized trial analysis given in Section 2.3, assuming an exponential sampling model for the observed recurrence times, as in (7) and focusing on only one treatment, say $j = 2$. In contrast with (7), we complete the model with conjugate Gamma and inverse Gamma priors and hyperpriors:

$$\begin{array}{ll}
 \text{Sampling model} & Y_{k,1}, \dots, Y_{k,m} \mid \theta_k \sim \text{Exp}(\theta_k) \text{ indep. for all } k \\
 \text{Priors} & \theta_k \mid \tilde{\mu}, \tilde{\gamma} \sim \text{Ga}(\tilde{\gamma}, \tilde{\gamma}/\tilde{\mu}) \text{ i.i.d. for all } k \\
 \text{Hyperpriors} & \tilde{\mu} \mid \alpha_\phi, \beta_\phi \sim \text{IG}(\alpha_\phi, \beta_\phi) \\
 & \tilde{\gamma} \mid a_\phi, b_\phi \sim \text{Ga}(a_\phi, b_\phi).
 \end{array}$$

The $\text{Ga}(\tilde{\gamma}, \tilde{\gamma}/\tilde{\mu})$ prior has mean $\tilde{\mu}$ and variance $\tilde{\mu}^2/\tilde{\gamma}$. We will use $\boldsymbol{\phi}_{\tilde{\mu}} = (\alpha_\phi, \beta_\phi)$ and $\boldsymbol{\phi}_{\tilde{\gamma}} = (a_\phi, b_\phi)$ to denote the hyperparameters specific to the population mean and the shape parameter, respectively. For case 1, where the goal is inference on $\theta_1, \dots, \theta_K$, we derive the ESS both with and without fixed shape parameter $\tilde{\gamma}$ at the second level.

Case 1, ESS with fixed $\tilde{\gamma}$: Assuming fixed $\tilde{\gamma}$, and using an approximation, the evaluation of Δ_1 is possible in closed form as follows. The target prior is

$$\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \tilde{\gamma}) = \left\{ \frac{\tilde{\gamma}^{\tilde{\gamma}}}{\Gamma(\tilde{\gamma})} \right\}^K \frac{\beta_\phi^{\alpha_\phi}}{\Gamma(\alpha_\phi)} (\theta_1, \dots, \theta_K)^{\tilde{\gamma}-1} \frac{\Gamma(K\tilde{\gamma} + \alpha_\phi)}{\left(\tilde{\gamma} \sum_{k=1}^K \theta_k + \beta_\phi \right)^{K\tilde{\gamma} + \alpha_\phi}},$$

where $\Gamma(\cdot)$ denotes a gamma function. Specifying the ε -information prior $\pi_{12,0}$ with $\pi_{1,0}(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}, c)$, $\boldsymbol{\theta} \mid \tilde{\mu}, \tilde{\gamma}, c \sim \prod_{k=1}^K \text{Ga}(\theta_k \mid \frac{\tilde{\gamma}}{c}, \frac{\tilde{\gamma}}{c\tilde{\mu}})$, one may compute

$$-H_{\pi_{12}}(\bar{\boldsymbol{\theta}}_{12}) = \frac{\tilde{\gamma} - 1}{(\bar{\theta}_{12})^2} \mathbf{I} - \frac{(K\tilde{\gamma} + \alpha_\phi)\tilde{\gamma}^2}{(K\tilde{\gamma} + \alpha_\phi - 1)^2(\bar{\theta}_{12})^2} \mathbf{J} \quad \text{and} \quad -H_{\pi_{12,M}}(\bar{\boldsymbol{\theta}}_{12}, \bar{\mathcal{Y}}_M) \approx \frac{m - 1}{(\bar{\theta}_{12})^2} \mathbf{I},$$

where $m = M/K$ and $\bar{\boldsymbol{\theta}}_{12}$ denotes a vector with entries the prior mean of $\theta_1, \dots, \theta_K$ under π_{12} , which is $\bar{\theta}_{12} = \int \int \theta_k \pi_1(\theta_k \mid \tilde{\mu}, \tilde{\gamma}) \pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}}) d\theta_k d\tilde{\mu} = \int \tilde{\mu} \pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}}) d\tilde{\mu} = \beta_\phi / (\alpha_\phi - 1)$. Note that the approximation for $-H_{\pi_{12,M}}(\bar{\boldsymbol{\theta}}_{12}, \bar{\mathcal{Y}}_M)$ is valid because c is large. Computing $\Delta_1(M, \pi_1, \pi_2, \pi_{1,0})$ gives the $\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}})$.

The left-hand half of Table S1 summarizes the ESS values computed for $\tilde{\gamma} = 20, 5$, and 2 and several fixed values of α_ϕ and β_ϕ , when $K = 5$. In the context of a non-hierarchical model, that is, under a gamma-exponential model with fixed values of $\tilde{\mu}$ and $\tilde{\gamma}$, the ESS of a gamma prior is $\tilde{\gamma}$ (MTM, Section 5, Example 2). Thus, for $K = 5$ independent priors the ESS should be 100. Table S1 shows how the values of $\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}})$ adjust for the dependence in the hierarchical model.

We evaluate $\rho = \text{corr}(\theta_k, \theta_{k'} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \tilde{\gamma})$ numerically in this example by sampling $(\theta_k, \theta_{k'})$ pairs from π_{12} . Given $\tilde{\gamma}$, as $\rho \rightarrow 0$ the K subgroups become independent and $\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}) \rightarrow K \times \tilde{\gamma}$, shown numerically in the fourth row of Table S1. As expected, the ESS becomes equivalent to the sum of the ESS values for K independent models. The ESS values are determined by α_ϕ , $\tilde{\gamma}$, and K , not β_ϕ . In contrast, as $\rho \rightarrow 1$, the ESS becomes smaller, also shown by Table S1. Because $\bar{\boldsymbol{\theta}}_{12}$ is canceled out in computing $\Delta_1(M, \pi_1, \pi_2, \pi_{1,0})$, β_ϕ does not affect the ESS in this example.

Case 1, ESS with random $\tilde{\gamma}$: When $\tilde{\gamma}$ is random, it is no longer possible to obtain closed form expressions for $\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}})$, and we apply the numerical methods described above. Similarly, one may obtain the ε -information prior $\pi_{12,0}$ with $\pi_{1,0}(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}, c)$, $\boldsymbol{\theta} \mid \tilde{\mu}, \tilde{\gamma}, c \sim \prod_{k=1}^K \text{Ga}(\theta_k \mid \frac{\tilde{\gamma}}{c}, \frac{\tilde{\gamma}}{c\tilde{\mu}})$. The prior mean $\bar{\boldsymbol{\theta}}_{12}$, may be specified analytically as a vector with

K elements $\bar{\theta}_{12} = \beta_\phi / (\alpha_\phi - 1)$.

The right hand side of Table S1 shows the resulting ESS values for several choices of $\phi_{\tilde{\mu}}$ and $\phi_{\tilde{\gamma}}$ when $K=5$. We use the same values for $\phi_{\tilde{\mu}} = (\alpha_\phi, \beta_\phi)$ as those used in the previous section, and $\phi_{\tilde{\gamma}} = (a_\phi, b_\phi)$ is set equal to (200,10), (50,10), and (20,10) so that $E_{\pi_2}(\tilde{\gamma} | \phi_{\tilde{\gamma}}) = 20, 5, \text{ and } 2$, respectively, which are the same as the fixed values of $\tilde{\gamma}$ used in the previous section. Here, $E_{\pi_2}(\tilde{\gamma} | \phi_{\tilde{\gamma}})$ denotes the prior mean of $\tilde{\gamma}$ under $\pi_2(\tilde{\gamma} | \phi_{\tilde{\gamma}})$. ESS values are obtained similarly as in the previous section. This is reasonable, given the similarity of the formulas (S1) and (S2) found in the normal/inverse χ^2 -normal-normal example.

Case 2: Using Algorithm 2 given in Section 3.3, we compute $\Delta_2(M, \pi_2, \pi_{2,0})$ first by specifying the ε -information priors as $\pi_{2,0}(\tilde{\mu} | \phi_{\tilde{\mu}}, c), \tilde{\mu} | \alpha_\phi, \beta_\phi, c \sim \text{IG}(2 + c^{-1}, \frac{\beta_\phi}{\alpha_\phi - 1})$, and $\pi_{2,0}(\tilde{\gamma} | \phi_{\tilde{\gamma}}, c), \tilde{\gamma} | a_\phi, b_\phi, c \sim \text{Ga}(\frac{a_\phi}{c}, \frac{b_\phi}{c})$. The diagonal (1,1) and (2,2) elements of the information matrix of $\pi_2(\tilde{\theta} | \phi)$, which respectively correspond to $\pi_2(\tilde{\mu} | \phi_{\tilde{\mu}})$ and $\pi_2(\tilde{\gamma} | \phi_{\tilde{\gamma}})$, are specified as $-H_{\pi_2}(\tilde{\theta})^{(1,1)} = -\frac{(\alpha_\phi + 1)}{\tilde{\mu}^2} + \frac{2\beta_\phi}{\tilde{\mu}^3}$ and $-H_{\pi_2}(\tilde{\theta})^{(2,2)} = \frac{(\alpha_\phi - 1)}{\tilde{\gamma}^2}$. The diagonal elements of the information matrix of $\pi_{2,M}(\tilde{\theta} | \phi_0, \bar{\mathcal{Y}}_{M,2})$ are found to be

$$-H_{\pi_{2,M}}(\tilde{\theta}, \bar{\mathcal{Y}}_{M,2})^{(1,1)} = -\frac{3 + c^{-1}}{\tilde{\mu}^2} + \frac{2\beta_\phi}{(\alpha_\phi - 1)\tilde{\mu}^3} - K\frac{\tilde{\gamma}}{\tilde{\mu}^2} + \sum_{k=1}^K \frac{\tilde{\gamma}(m + \tilde{\gamma})(2\tilde{\mu}m\bar{Y}_2 + \tilde{\gamma})}{\{\tilde{\mu}^2(m\bar{Y}_2 + \tilde{\gamma}/\tilde{\mu})\}^2}$$

and

$$\begin{aligned} -H_{\pi_{2,M}}(\tilde{\theta}, \bar{\mathcal{Y}}_{M,2})^{(2,2)} &= \frac{(\alpha_\phi/c - 1)}{\tilde{\gamma}^2} - \frac{K}{\tilde{\gamma}} + K \sum_{i=1}^m \frac{1}{(\tilde{\gamma} + i - 1)^2} \\ &+ \sum_{k=1}^K \left\{ \frac{1}{\tilde{\mu}m\bar{Y}_2 + \tilde{\gamma}} + \frac{\tilde{\mu}m\bar{Y}_2 - m}{(\tilde{\mu}m\bar{Y}_2 + \tilde{\gamma})^2} \right\}, \end{aligned}$$

where $\bar{Y}_2 = E(Y_{k,i} | \tilde{\theta}_2)$. The marginal likelihood $f_1(\mathcal{Y}_M | \tilde{\mu}, \tilde{\gamma})$ is analytically specified as

$$f_1(\mathcal{Y}_M | \tilde{\mu}, \tilde{\gamma}) = \frac{\Gamma(m + \tilde{\gamma})^K}{\prod_{k=1}^K (\sum_{i=1}^m Y_{k,i} + \tilde{\gamma}/\tilde{\mu})^{m+\tilde{\gamma}}} \left\{ \frac{(\tilde{\gamma}/\tilde{\mu})^{\tilde{\gamma}}}{\Gamma(\tilde{\gamma})} \right\}^K.$$

Identifying $\bar{\boldsymbol{\theta}}_2 = \mathbb{E}_{\pi_2}(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{\phi}) = (\frac{\beta_\phi}{\alpha_\phi - 1}, \frac{a_\phi}{b_\phi})$, then, computing $\Delta_2(M, \pi_2, \pi_{2,0})$, $\text{ESS}_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}})$ and $\text{ESS}_2(\tilde{\gamma} \mid \boldsymbol{\phi}_{\tilde{\gamma}})$ may be obtained.

When the goal is inference on the population mean parameter, the ESS is computed for the hyperprior, $\pi_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}})$, of the population mean. The results are summarized in Table S2 for several choices of α_ϕ and β_ϕ with $a_\phi = 5$ and $b_\phi = 1$. Since the marginal variance of $\tilde{\mu}$ is $\beta_\phi^2/(\alpha_\phi - 1)^2(\alpha_\phi - 2)$, it follows that $\text{ESS}_2(\tilde{\mu} \mid \boldsymbol{\phi}_{\tilde{\mu}}) \rightarrow 0$ as $\alpha_\phi \rightarrow 2$.

3. Beta/Gamma-Beta-Binomial Model

As a final example, we consider a hierarchical model for binary responses recorded over different subpopulations, as in the sarcoma trial of Section 2.1. We assume a binomial sampling model with sub-population specific success probabilities θ_k . In contrast with (2.1), we complete the model with a conjugate Beta prior on the untransformed probabilities $\theta_1, \dots, \theta_K$, and beta/gamma hyperpriors:

$$\begin{aligned} \text{Sampling model} \quad Y_{k,m_k} \mid \theta_k &\sim \text{Bin}(m_k, \theta_k) \text{ indep. for all } k \\ \text{Prior} \quad \theta_k \mid \tilde{\xi}, \tilde{\gamma} &\sim \text{Be}(\tilde{\gamma}\tilde{\xi}, \tilde{\gamma}(1 - \tilde{\xi})) \text{ i.i.d. for all } k \\ \text{Hyperpriors} \quad \tilde{\xi} \mid \alpha_\phi, \beta_\phi &\sim \text{Be}(\alpha_\phi, \beta_\phi) \\ &\tilde{\gamma} \mid a_\phi, b_\phi \sim \text{Ga}(a_\phi, b_\phi). \end{aligned}$$

The beta prior is parameterized to have mean $\tilde{\xi}$ and variance $\tilde{\xi}(1 - \tilde{\xi})/(\tilde{\gamma} + 1)$. For case 1, we compute the ESS for $\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_\xi, \boldsymbol{\phi}_\gamma)$ where $\boldsymbol{\phi}_\xi = (\alpha_\phi, \beta_\phi)$, $\boldsymbol{\phi}_\gamma = (a_\phi, b_\phi)$, and for case 2 we compute $\text{ESS}_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$ and $\text{ESS}_2(\tilde{\gamma} \mid \boldsymbol{\phi}_\gamma)$.

Case 1: Because a closed form of $\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_\xi, \boldsymbol{\phi}_\gamma)$ cannot be obtained, we compute $-H_{\pi_{12}}(\bar{\boldsymbol{\theta}}_{12})$ using the simulation-based approximation. We also obtain the ε -information prior $\pi_{12,0}$ for $\pi_{1,0}(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}, c)$, $\boldsymbol{\theta} \mid \tilde{\xi}, \tilde{\gamma}, c \sim \prod_{k=1}^K \text{Be}(\theta_k \mid \frac{\tilde{\gamma}\tilde{\xi}}{c}, \frac{\tilde{\gamma}(1-\tilde{\xi})}{c})$. The prior mean $\bar{\boldsymbol{\theta}}_{12}$ is a K -vector with all elements $\bar{\theta}_{12} = \frac{\alpha_\phi}{\alpha_\phi + \beta_\phi}$, and the information matrix of $f(\mathcal{Y}_M \mid \boldsymbol{\theta})$ is a

$K \times K$ matrix with k^{th} diagonal element $\frac{Y_{k,m}}{\theta_k^2} + \frac{m-Y_{k,m}}{(1-\theta_k)^2}$ and all off-diagonal elements 0.

Table S3 gives values of $\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi})$ for $K = 5$ subpopulations and several choices of α_ϕ, β_ϕ . We set $(\alpha_\phi, \beta_\phi)$ so that $E_{\pi_2}(\tilde{\xi} \mid \alpha_\phi, \beta_\phi) = 2/3$, and we set (a_ϕ, b_ϕ) so that $E_{\pi_2}(\tilde{\gamma} \mid a_\phi, b_\phi) = 20$ or 5. Under a non-hierarchical Binomial/Beta model the prior ESS is commonly reported as the sum of the Beta parameters, which in our case equals $\tilde{\gamma}$. Thus, the ESS for K independent subpopulations is $K\tilde{\gamma}$. Note that, because $\tilde{\gamma}$ is random in the above model, the prior mean of $\tilde{\gamma}$ under $\pi_2(\tilde{\gamma} \mid a_\phi, b_\phi)$, $E_{\pi_2}(\tilde{\gamma} \mid a_\phi, b_\phi)$ should be used here rather than $\tilde{\gamma}$. Table S3 shows the adjustment for dependence, which substantially reduces $K \times E_{\pi_2}(\tilde{\gamma} \mid a_\phi, b_\phi)$. For example, if the above model were used for a clinical trial similar to the sarcoma trial from Section 2.1, and the trial enrolled 100 patients, any of the first six prior choices in Table S3 would be unreasonably influential.

The plausibility of the reported prior ESS values in Table S3 may be evaluated by considering a limiting case. Let $\rho = \text{corr}(\theta_k, \theta_{k'})$ under π_{12} . As $\rho \rightarrow 0$, we expect the prior ESS to match the prior ESS of $K = 5$ separate non-hierarchical Beta/Binomial models, each with prior ESS equal to the sum of the Beta parameters. This is verified by noting that, for smaller ρ , $\text{ESS}_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi})$ is closer to $K \times E_{\pi_2}(\tilde{\gamma} \mid a_\phi, b_\phi) = 5 \times 20 = 100$ in rows 1 - 4 and $5 \times 5 = 25$ in rows 5 - 8 in Table S3.

Case 2: ESS for the hyperprior on $\tilde{\xi}$: Using Algorithm 2 given in Section 3.3, we compute $\Delta_2(M, \pi_2, \pi_{2,0})$. The ε -information priors are specified as $\pi_{2,0}(\tilde{\xi} \mid \boldsymbol{\phi}_{\tilde{\xi}}, c)$, $\tilde{\xi} \mid \alpha_\phi, \beta_\phi, c \sim \text{Be}(\frac{\alpha_\phi}{c}, \frac{\beta_\phi}{c})$, and $\pi_{2,0}(\tilde{\gamma} \mid \boldsymbol{\phi}_{\tilde{\gamma}}, c)$, $\tilde{\gamma} \mid a_\phi, b_\phi, c \sim \text{Ga}(\frac{a_\phi}{c}, \frac{b_\phi}{c})$. The diagonal (1,1) and (2,2) elements of the information matrix of $\pi_2(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{\phi})$, which respectively correspond to $\pi_2(\tilde{\xi} \mid \boldsymbol{\phi}_{\tilde{\xi}})$ and $\pi_2(\tilde{\gamma} \mid \boldsymbol{\phi}_{\tilde{\gamma}})$, are $-H_{\pi_2}(\tilde{\boldsymbol{\theta}})^{(1,1)} = \frac{\alpha_\phi - 1}{\tilde{\xi}^2} + \frac{\beta_\phi - 1}{(1 - \tilde{\xi})^2}$ and $-H_{\pi_2}(\tilde{\boldsymbol{\theta}})^{(2,2)} = \frac{\alpha_\phi - 1}{\tilde{\gamma}^2}$, and the marginal likelihood is

$$f_1(\mathcal{Y}_M \mid \tilde{\xi}, \tilde{\gamma}) = \prod_{k=1}^K \binom{m}{Y_{k,m}} \frac{\prod_{i'=1}^{Y_{k,m}} (\tilde{\gamma}\tilde{\xi} + i' - 1) \prod_{i''=1}^{m-Y_{k,m}} (\tilde{\gamma}(1 - \tilde{\xi}) + i'' - 1)}{\prod_{i=1}^m (\tilde{\gamma} + i - 1)}.$$

The diagonal elements of the information matrix of $\pi_{2,M}(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{\phi}_0, \bar{\mathcal{Y}}_{M,2})$ are found to be

$$-H_{\pi_{2,M}}(\tilde{\boldsymbol{\theta}}, \bar{\mathcal{Y}}_{M,2})^{(1,1)} = \frac{\frac{\alpha_\phi}{c} - 1}{\tilde{\xi}^2} + \frac{\frac{\beta_\phi}{c} - 1}{(1 - \tilde{\xi})^2} + \sum_{k=1}^K \left\{ \sum_{i'=1}^{\bar{Y}_{m,2}} \frac{\tilde{\gamma}^2}{(\tilde{\gamma}\tilde{\xi} + i' - 1)^2} + \sum_{i''=1}^{m-\bar{Y}_{m,2}} \frac{\tilde{\gamma}^2}{\{\tilde{\gamma}(1 - \tilde{\xi}) + i'' - 1\}^2} \right\}$$

and

$$-H_{\pi_{2,M}}(\tilde{\boldsymbol{\theta}}, \bar{\mathcal{Y}}_{M,2})^{(2,2)} = \frac{\frac{\alpha_\phi}{c} - 1}{\tilde{\gamma}^2} + \sum_{k=1}^K \left\{ \sum_{i'=1}^{\bar{Y}_{m,2}} \frac{\tilde{\xi}^2}{(\tilde{\gamma}\tilde{\xi} + i' - 1)^2} + \sum_{i''=1}^{m-\bar{Y}_{m,2}} \frac{(1 - \tilde{\xi})^2}{\{\tilde{\gamma}(1 - \tilde{\xi}) + i'' - 1\}^2} - \sum_{i=1}^m \frac{1}{(\tilde{\gamma} + i - 1)^2} \right\},$$

where $\bar{Y}_{m,2} = E(Y_{k,m} \mid \bar{\boldsymbol{\theta}}_2)$. Given the prior means of the hyperparameters of $\bar{\boldsymbol{\theta}}_2 = E_{\pi_2}(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{\phi}) = (\frac{\alpha_\phi}{\alpha_\phi + \beta_\phi}, \frac{a_\phi}{b_\phi})$, by computing $\Delta_2(M, \pi_2, \pi_{2,0})$, ESSs of $\pi_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$ and $\pi_2(\tilde{\gamma} \mid \boldsymbol{\phi}_\gamma)$ may be obtained.

In Table S4, we summarize $\text{ESS}_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$ for several choices of α_ϕ and β_ϕ with $a_\phi = 20$ and $b_\phi = 10$. We fix $(\alpha_\phi, \beta_\phi)$ such that $E_{\pi_2}(\tilde{\xi} \mid \alpha_\phi, \beta_\phi) = 5/6$. Here, $\pi_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$, $\tilde{\xi} \mid \alpha_\phi, \beta_\phi \sim \text{Be}(\alpha_\phi, \beta_\phi)$, may represent prior information on an overall effect for all K subgroups, because $E_{\pi_1}(\theta_k \mid \tilde{\xi}, \tilde{\gamma}) = \tilde{\xi}$ for $k = 1, \dots, K$. Since the variance of $\tilde{\xi}$ under $\pi_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$ is $E_{\pi_2}(\tilde{\xi} \mid \alpha_\phi, \beta_\phi) \left\{ 1 - E_{\pi_2}(\tilde{\xi} \mid \alpha_\phi, \beta_\phi) \right\} / (\alpha_\phi + \beta_\phi + 1)$, the ESS of $\pi_2(\tilde{\xi} \mid \boldsymbol{\phi}_\xi)$ increases with $\alpha_\phi + \beta_\phi$.

Table S1. $ESS_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi})$ values in case 1 for the gamma/inverse gamma-gamma-exponential model, with and without fixed shape parameter $\tilde{\gamma}$ at the second level, with $\boldsymbol{\phi}_{\tilde{\mu}} = (\alpha_{\phi}, \beta_{\phi})$ and $\boldsymbol{\phi}_{\tilde{\gamma}} = (a_{\phi}, b_{\phi})$, for $K = 5$ subgroups. When $\tilde{\gamma}$ is not fixed, $\hat{\tilde{\gamma}} = E_{\pi_2}(\tilde{\gamma} \mid a_{\phi}, b_{\phi})$.

$\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \tilde{\gamma}), \text{ for } \tilde{\gamma} \text{ fixed}$					$\pi_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\mu}}, \boldsymbol{\phi}_{\tilde{\gamma}}), \text{ for } \tilde{\gamma} \text{ not fixed}$						
$\tilde{\gamma}$	α_{ϕ}	β_{ϕ}	ρ	ESS ₁₂	$\hat{\tilde{\gamma}}$	α_{ϕ}	β_{ϕ}	a_{ϕ}	b_{ϕ}	ρ	ESS ₁₂
20	100	100	0.168	86.6	20	100	100	200	10	0.168	87.3
20	50	50	0.290	79.1	20	50	50	200	10	0.289	79.9
20	15	15	0.586	60.6	20	15	15	200	10	0.587	61.4
5	100	100	0.048	23.9	5	100	100	50	10	0.046	24.9
5	50	50	0.092	22.9	5	50	50	50	10	0.091	23.4
5	15	15	0.263	19.2	5	15	15	50	10	0.262	19.7
2	20	20	0.094	8.9	2	20	20	20	10	0.090	9.5
2	10	10	0.181	4.5	2	10	10	20	10	0.174	4.1
2	6	6	0.281	2.9	2	6	6	20	10	0.278	0.3

Table S2. $ESS_2(\tilde{\mu}|\alpha_\phi, \beta_\phi)$ values in case 2 for the gamma/inverse gamma-gamma-exponential model, for $K = 5$ subgroups, where a_ϕ and b_ϕ denote the hyperparameters of the hyperprior $\pi_2(\tilde{\gamma} | a_\phi, b_\phi)$.

α_ϕ	β_ϕ	a_ϕ	b_ϕ	ESS_2
20	3	5	1	95.1
10	3	5	1	16.9
2.1	3	5	1	0.3
2.01	3	5	1	0.03
2.001	3	5	1	0.003

Table S3. $ESS_{12}(\boldsymbol{\theta} \mid \boldsymbol{\phi}_{\tilde{\xi}}, \boldsymbol{\phi}_{\tilde{\gamma}})$ values in case 1 for the beta/gamma-beta-binomial model, for $K = 5$. The mean values $E_{\pi_2}(\tilde{\gamma} \mid \boldsymbol{\phi}_{\tilde{\gamma}}) = 20$ and 5 are obtained by setting the parameters of the hyperprior $\pi_2(\tilde{\gamma} \mid \boldsymbol{\phi}_{\tilde{\gamma}})$ to $(\mathbf{a}_\phi, \mathbf{b}_\phi) = (200, 10)$ and $(50, 10)$, respectively.

$E_{\pi_2}(\tilde{\gamma} \mid \mathbf{a}_\phi, \mathbf{b}_\phi)$	α_ϕ	β_ϕ	ρ	ESS_{12}
20	40	20	0.259	82.3
20	20	10	0.412	75.7
20	10	5	0.582	62.5
20	6	3	0.700	51.8
5	40	20	0.090	24.4
5	20	10	0.165	21.8
5	10	5	0.283	8.4
5	6	3	0.396	3.5

Table S4. $ESS_2(\tilde{\xi} \mid \alpha_\phi, \beta_\phi)$ values in case 2 for the beta/gamma-beta-binomial model, for $K = 5$, where a_ϕ and b_ϕ are the hyperparameters of $\pi_2(\tilde{\gamma} \mid a_\phi, b_\phi)$.

α_ϕ	β_ϕ	a_ϕ	b_ϕ	ESS_2
50	10	20	10	25.2
20	4	20	10	4.8
5	1	20	10	1.2
2.5	0.5	20	10	0.6
1.5	0.3	20	10	0.4