

Effective Sample Size Calculator for Basket trials

(Ver 1.0)

Xin Chen

Contents

1 Overview.....	2
2 Module	2
2.1 JAGS Installation	2
2.2 Data Analysis	2
2.2.1 Type of BHM.....	2
2.2.2 Hyperprior	3
2.2.3 MCMC.....	3
2.2.4 Data input	3
2.2.5 Output	4
2.3 Convergence diagnostics.....	4
2.4 Calculating ESS.....	4
2.4.1 ESS based on MSE	4
2.4.2 ESS based on variance ratio	5
2.4.3 ESS based on moment matching	5

1 Overview

This software was developed with R shiny and deployed as a Windows desktop application. It facilitates fitting Bayesian hierarchical model (BHM) in basket trials, and calculations of effective sample size (ESS) in a way that does not require programming. All output figures and tables can be downloaded.

The current version of the software only supports binary endpoints, but the corresponding R code has been attached with the software, allowing statisticians to customize it for specific trial designs.

2 Module

2.1 JAGS Installation

The software facilitates fitting BHM by JAGS (Just Another Gibbs Sampler), which needs to be installed on the local computer before further analysis. If it is not installed yet, click the button ‘Install JAGS’ and follow the prompts to install it. Then, type the installation path of the JAGS in the input box and click the button ‘Interface to the JAGS’ to connect the software with JAGS. After that, click the button ‘Check the status of JAGS installation’ to see whether the JAGS is ready for use.

There is no need to repeat this step after the first use.

2.2 Data Analysis

2.2.1 Type of BHM

In the panel on the left, the users can define the model settings of BHM. We support two types of BHM here, one is borrowing information across cancer types on the log odds scale (Type 1); the other is borrowing on the log odds ratio scale (Type 2). Consider a single-arm basket trial designed to evaluate the treatment effects of the investigational therapy for K indications. Let $Y_k (k = 1, \dots, K)$ denote the number of responders among n_k enrolled patients with indication k , and p_k denotes the probability of response for indication k . Then a BHM of Type 1 is as follows:

$$Y_k \sim \text{Bin}(n_k, p_k)$$

$$\theta_k = \text{logit}(p_k)$$

$$\theta_k \sim N(\mu, \sigma^2).$$

The BHM of Type 2 is as follows:

$$\begin{aligned} Y_k &\sim \text{Bin}(n_k, p_k) \\ \theta_k &= \text{logit}(p_k) - \text{logit}(p_{k0}) \\ \theta_k &\sim N(\mu, \sigma^2), \end{aligned}$$

where p_{k0} is the pre-specified null response rate (usually the response rate of the standard of care) for the k_{th} indication.

2.2.2 Hyperprior

Some common choices of hyperpriors are provided here. For the average treatment μ , a normal hyperprior distribution is set, ie, $\mu \sim N(\mu_0, \sigma_0^2)$, where μ_0 is the prior mean and σ_0^2 is the prior variance.

In terms of the between-group variance σ^2 , there are several choices for the hyperprior:

- (1) Inverse gamma distribution for the variance parameter: $\sigma^2 \sim IG(a, b)$.
- (2) Half-normal distribution for the scale parameter: $\sigma \sim HN(0, \tau^2)$, where the support is $\sigma \geq 0$.
- (3) Half- t distribution for the scale parameter: $\sigma \sim Ht(0, \tau, df)$, where the support is $\sigma \geq 0$.
- (4) Uniform distribution for the scale parameter: $\sigma \sim Unif(a, b)$.

In the panel on the right, the prior densities for μ , σ and σ^2 are plotted in real time.

2.2.3 MCMC

The fitting of BHM is based on the Markov Chain Monte Carlo (MCMC) approach. Options regarding MCMC include the number of MCMC chains, number of iterations for adaptation, number of iterations for sampling, and thinning interval.

2.2.4 Data input

Data can be uploaded into the software in the format of Excel (.xlsx). A file is attached as an example, named 'Data example.xlsx'. Note that there must be columns with names 'Indication', 'n', and 'y' in the file. If the BHM type is selected as Type 2, the Excel file also needs to have a column named 'p0'. 'n', 'y', and 'p0' denote the sample size, the number of responders, and the null response rate for each indication.

2.2.5 Output

After model setup and data entry, click the button ‘Run!’ in the left panel to conduct data analysis. The results will then be exhibited in the subpanel of ‘Result’ and ‘Forest Plot’ on the right. The frequentist estimate is the observed ORR with a 95% confidence interval (CI) based on Clopper and Pearson method. The BHM estimate is the posterior mean with a 95% equal-tailed credible interval (CrI).

2.3 Convergence diagnostics

Several diagnostic tools for MCMC convergence are available in this software, including the posterior density, traces of MCMC chains, autocorrelation plot, effective sample size for MCMC sampling, and Gelman-Rubin diagnostic plot. In the left panel, the users can select the parameters of interest. Then, the corresponding results for convergence diagnostics will be displayed in the right panel.

2.4 Calculating ESS

Calculating ESS may be time-consuming, especially when the number of indications is large. We have therefore designed parallel computing options for users. In the left panel, the users will see a piece of prompt text in red, showing the number of CPU cores inside the computer. Then, enter the number of CPU cores we intend to use (it is recommended not to use more than half of all cores, otherwise the computer may crash), and click the button ‘Run!’. After waiting for some time, the results will shown in the right panel.

Three types of prior ESS are supported in the software, and we will next introduce them one by one.

2.4.1 ESS based on MSE

Let y_k denote the observed number of responders in indication k . For ESS based on mean squared error (MSE), the true value of p_k is unknown and can be replaced by the maximum likelihood estimate (MLE) $\hat{p}_k = y_k/n_k$ when deriving the ESS. Then, the difference of MSE for indication k between the independent analysis with a sample of size $m_k + n_k$ and the BHM with a sample of size n_k plus data from other indications is given by

$$E_{Y_{k\{m_k+n_k\}}} \left\{ SE_{Ind}^{\hat{p}_k}(p_k | Y_{k\{m_k+n_k\}}, m_k + n_k) \right\} - E_{Y_{k\{n_k\}}} \left\{ SE_{BHM}^{\hat{p}_k}(p_k | Y_{k\{n_k\}}, n_k, y_{-k}, n_{-k}) \right\}.$$

$SE_{Ind}^{\hat{p}_k}(p_k | Y_{k\{m_k+n_k\}}, m_k + n_k)$ denotes the squared error (SE) for the independent analysis with a sample size of $m_k + n_k$, and is defined as $(Y_{k\{m_k+n_k\}}/(m_k + n_k) - \hat{p}_k)^2$. The implication of $SE_{BHM}^{\hat{p}_k}(p_k | Y_{k\{n_k\}}, y_{-k}, n_{-k})$ is $(\hat{p}_{k,BHM} - \hat{p}_k)^2$, where

$\hat{p}_{k,BHM}$ is the posterior mean for p_k by fitting the BHM conditional on $Y_{k\{n_k\}}$ and data from indications other than k (ie, y_{-k} and n_{-k}). The MSE is then obtained by further averaging SE on the predictive distribution of $Y_{k\{m_k+n_k\}}$ (or $Y_{k\{n_k\}}$), which is assumed to be a binomial distribution with probability \hat{p}_k and size of $m_k + n_k$ (or n_k). The prior ESS for indication k is then given by m_k that minimizes the difference of MSE. By selecting the indications of interest, the users can see the process of searching m_k in the ‘Details’ subpanel on the right.

2.4.2 ESS based on variance ratio

The prior ESS based on variance ratio is defined as

$$\frac{Var(p_k|Independent\ analysis)}{Var(p_k|BHM)}(n_k - 1),$$

where the primary concern is the reduction of variance on the estimators. $Var(p_k|Independent\ analysis)$ is obtained by $\hat{p}_k(1 - \hat{p}_k)/n_k$.

2.4.3 ESS based on moment matching

The prior ESS based on moment matching is calculated as follows. After fitting the BHM, the posterior mean and posterior variance for each p_k are obtained, denoted as $E(p_k|BHM)$ and $Var(p_k|BHM)$ respectively. Then, it is not difficult to find a Beta distribution $Beta(a_k, b_k)$ with a mean of $E(p_k|BHM)$ and a variance of $Var(p_k|BHM)$. The prior ESS based on moment matching is then $a_k + b_k$ by the convention in the case of conjugate priors.