MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Bayesian Methods for Clinical Trials

by Libby Daniells & Pavel Mozgunov & Thomas Jaki

24th – 25th of February 2025

Practical 4: Borrowing of information in basket trials

Recall the BRAF-V600 basket trial in Lecture 9. It is a phase II PoC trial to smultaneously evaluate the efficacy of vemurafenib, a novel anti-cancer treatment, in several patient subpopulations. The primary endpoint is binary: response or non-response.
Open the R script titled `BasketTrials.R` and ensure that the JAGS model files, `StandardHM.txt`, `EXNEX.txt` and `Stand-alone.txt`, are in the working directory (Use `setwd()` to modify if necessary). For each substudy, the Bayesian analysis model returns the posterior mean of response rate together with a credible interval.

(a) Read the `StandardHM.txt` and `EXNEX.txt` scripts to understand how these Bayesian models are fitted for borrowing of information.

(b) Use the code written in `BasketTrials.R` to run an analysis for the BRAF-V600 trial.

```
# BRAF-V600 basket trial data
> nMod = 6
> r = c(8, 1, 6, 0, 2, 6)
> n = c(20, 8, 18, 10, 7, 32)
```

Set $\mu \sim N(0, 10^2)$, $\tau^2 \sim HN(0.125)$ to implement the standard hierarchical model and, additionally, the prior mixture weight as (0.5, 0.5) to implement the EXNEX model. For comparison, implement the stand-alone analysis model.

What are the posterior mean for the response rate, denoted by p_k, per substudy $k = 1, \ldots, 6$? What are the posterior probabilities that $\Pr(p_k > 0.25 \mid \text{data})$?

Method	Posterior Mean Response Rates						$\mathbb{P}(p_k > 0.25\|\text{data})$					
Basket	1	2	3	4	5	6	1	2	3	4	5	6
Independent	0.400	0.129	0.336	0.010	0.290	0.187	0.926	0.139	0.778	0.002	0.546	0.171
BHM	0.252	0.240	0.249	0.236	0.245	0.239	0.490	0.405	0.462	0.379	0.433	0.389
EXNEX	0.324	0.222	0.291	0.036	0.274	0.230	0.998	0.836	0.992	0.123	0.956	0.949

(c) Install and load the 'bmabasket' package in R. Using the bma() function, run an analysis for the BRAF-V600 trial using Bayesian model averaging. What are the posterior means of the response rates per substudy? What are the posterior probabilities that $\mathbb{P}(p_k > 0.25|\text{data})$?

```
library(bmabasket)
bma(0.25,y=r,n=n)

    $bmaProbs
           [,1]
[1,] 0.87465297
[2,] 0.28755094
[3,] 0.74199426
[4,] 0.04953094
[5,] 0.57953209
[6,] 0.28504852

$bmaMeans
          [,1]
[1,] 0.3604219
[2,] 0.1922144
[3,] 0.3172860
[4,] 0.0817006
[5,] 0.2845246
[6,] 0.2103263
```

$bmaProbs are the posterior probability of exceeding 0.25, $bmaMeans are the posterior means for each of the 6 sub-groups.

(d) Retain the prior specification unchanged from (b). Analyse the hypothetical datasets below using the standard hierarchical modelling approach, EXNEX model, BMA and stand-alone analyis. Again, what are the posterior mean for the response rate, denoted by p_k, per substudy $k = 1, \ldots, 5$? What are the posterior probabilities that $\Pr(p_k > 0.25 \mid \text{data})$?

Hypothetical dataset of a consistency scenario:

	Substudy 1	Substudy 2	Substudy 3	Substudy 4	Substudy 5
Number of patients	10	10	20	20	17
Number of responses	5	4	12	7	9

Hypothetical dataset of an inconsistency scenario:

	Substudy 1	Substudy 2	Substudy 3	Substudy 4	Substudy 5
Number of patients	10	10	20	20	17
Number of responses	2	7	15	4	5

Replace the BRAF-V600 basket trial data with the hypothetical datasets by stipulating:

```
# consistency scenario
> nMod = 5
```

```
> n = c(10, 10, 20, 20, 17)
> r = c( 5,  4, 12,  7,  9)
And
# inconsistency scenario
> nMod = 5
> n = c(10, 10, 20, 20, 15)
> r = c( 2,  7, 15,  4,  5)
```

Consistency scenario:

Method	Posterior Mean Response Rates					$\mathbb{P}(p_k > 0.25\|\text{data})$				
Basket	1	2	3	4	5	1	2	3	4	5
Independent	0.499	0.399	0.601	0.351	0.531	0.953	0.836	1.000	0.827	0.993
BHM	0.481	0.478	0.487	0.474	0.484	1.000	1.000	1.000	1.000	1.000
EXNEX	0.486	0.461	0.527	0.432	0.498	1.000	0.998	1.000	0.999	1.000
BMA	0.490	0.443	0.551	0.402	0.509	0.983	0.947	1.000	0.933	0.997

Inconsistency scenario:

Method	Posterior Mean Response Rates					$\mathbb{P}(p_k > 0.25\|\text{data})$				
Basket	1	2	3	4	5	1	2	3	4	5
Independent	0.200	0.700	0.749	0.200	0.334	0.302	0.999	1.000	0.265	0.752
BHM	0.422	0.458	0.479	0.409	0.429	0.990	1.000	1.000	0.985	0.993
EXNEX	0.268	0.614	0.707	0.246	0.339	0.894	1.000	1.000	0.915	0.989
BMA	0.244	0.684	0.727	0.231	0.314	0.433	1.000	1.000	0.386	0.676

(e) Compare the resulting posterior mean estimate and posterior standard deviation under both scenarios. What do you notice about the relative advantages and disadvantages of each analysis model?

Standard hierarchical modelling outperforms the EXNEX for analysing basket trials of a consistent scenario, while EXNEX is more robust to basket trials that have some extreme substudies. When an extreme substudy contains enough number of patients (i.e., borrowing of information is not desired), stand-alone analysis guarantees the accuracy of estimates.

(f) Set the argument of `Prior.tau.HN` to a larger value, for example, 0.25. What impact does this change have on the posterior means and credible intervals?

It leads to reduced borrowing of information, compared with results yielded by setting Prior.tau.HN = 0.125 under the same Bayesian analysis model for borrowing, i.e., StandardHM and EXNEX. Both approaches (especially the standard hierarchical modelling) tend to shrink the parameters towards their common mean to a smaller extent, which would therefore be more robust for analysing trial data of an inconsistent scenario.

(g) When using the EXNEX for analysis, set the argument of `pMix` to a range of large values, for example, 0.8, for the probability of exchangeability. How does this quantity affect the results?

The results suggest that EXNEX pushes more information to be shared across substudies. When there are substudies with low treatment effect, due to strong borrowing of information, it becomes more likely to declare the treatment is efficacious in that substudy.

(h) How would you expect the subtrial-wise type I error rate and statistical power to change using analysis models that permit borrowing of information? Perform a small-scale simulation study to develop intuition.

In inconsistent scenarios, when information borrowing is implemented we expect inflation in the type I error, with models with the strongest borrowing (e.g. BHM) demonstrating the greatest type I error rate. We also expect these methods to have the greatest power in consistent scenarios. A stand-alone analysis will protect the type I error rate.

Example simulation for the BHM

```
nMod <- 5
n <- rep(13,nMod)
p.1 <- rep(0.25,nMod) #Change this to run the simulation under different scenarios

samples.success <- matrix(NA,nrow=1000,ncol=nMod)
for(i in 1:1000){
  r <- rbinom(nMod,n,p.1)
  data_list <- list(
    nMod = nMod,
    r = r,
    n = n,
    Prior.mu.theta = c(0, 10),
    Prior.tau.HN = 0.125,
    p.cut = 0.25
  )
  inits <- list(mu.theta = 0)=
  iter <- 10000
  jags <- jags.model(file = "StandardHM.txt", data = data_list,
                     n.chains = 1,  n.adapt = iter)

  update(jags, iter, progress.bar = "none")
  mcmc.sampling <- jags.samples(jags, c("p", "success"), iter, progress.bar = "none")
  for(k in 1:nMod){
    samples.success[i,k] <- mean(mcmc.sampling$success[k,,])
  }
}
```

```
library(matrixStats)
Delta <- mean(colQuantiles(samples.success,probs=0.9)) #Calibrate the decision criteria
apply(samples.success>Delta,2,mean) #Compute the type I error rate
```

(i) Do you have any suggestions on how to better control for a type I error?

- Adjusting calibration to better control for a type I error.
- Data-driven approaches to make the models more sensitive to heterogeneity in the response rates across baskets.