Meta Analysis and Multiple Testing

Sujit K Ghosh

NC State University
DEPARTMENT OF STATISTICS

Web: http://www.stat.ncsu.edu/people/ghosh/
Email: sujit_ghosh@ncsu.edu

Presented at:
"Difficulties with Observational Medical Studies"
NISS Workshop, RTP, NC, USA

Meta-Analysis

- A quantitative synthesis of relevant information from multiple studies has become increasingly popular under the name of *Meta-Analysis*
- Often the goal is to provide a numerical summary of the overall effect of an intervention (e.g., treatment, device or service)
- Two models are frequently used in meta-analysis:
  - *Fixed effects model*: assumes that each observed study result is estimating a *common* overall effect
  - *Random effects model*: assumes that each observed result is estimating its *own* unknown effect; which in turn are estimating a common overall mean
- The use of either models depends on the context

- The fixed effects model does not allow for between-study heterogeneity
- The random effects model allows for between-study heterogeneity as well as within-study variability
- Consider K studies, each measuring the effect of an intervention
- For each $k = 1, 2, \ldots, K$, suppose
  - $\hat{\theta}_k$ denotes (estimated) effect within k-th study, and
  - s_k^2 denotes the associated variance within k-th study
- The fixed effects model assumes: $\hat{\theta}_k = \mu + s_k \epsilon_k$
  where μ is the overall effect, $E[\epsilon_k] = 0$ and $Var[\epsilon_k] = 1$
- In contrast, the random effects model assumes:
  $$\hat{\theta}_k = \theta_k + s_k \epsilon_k \quad \text{and} \quad \theta_k = \mu + \sigma e_k$$
  where θ_k's are (unknown) study-specific (random) effects and σ^2 is the between-study variance (assume $E[e_k] = 0$ and $Var[e_k] = 1$)

- Integrating the random effects we obtain: $\hat{\theta}_k = \mu + (\sigma e_k + s_k \epsilon_k)$
- Assuming $e_k \perp \epsilon_k$, we have $Var[\hat{\theta}_k] = s_k^2 + \sigma^2$
- Thus, under fixed effects model: $Var[\hat{\theta}_k] = s_k^2$
  whereas under random effects model: $Var[\hat{\theta}_k] = s_k^2 + \sigma^2$
- If study effects are assumed independent, then
  $$\hat{\mu} = \frac{\sum_{k=1}^K w_k \hat{\theta}_k}{\sum_{k=1}^K w_k} \quad \text{and} \quad Var[\hat{\mu}] = \frac{1}{\sum_{k=1}^K w_k}$$
  where the weights are given by $w_k = 1/Var[\hat{\theta}_k]$
- Allowing the extra between-studies variation (σ^2) has the effect of reducing the relative weights given to more precise studies
- Thus, random effects model produces more conservative interval estimates for the overall effect μ
- How do we estimate σ? *Should we treat s_k^2 as (known) fixed?*

- Assume that, under the random effects model
  (i) study effects, $\hat{\theta}_k \overset{ind}{\sim} N(\theta_k, s_k^2) \cdots\cdots (1)$
  (ii) random effects, $\theta_k \overset{iid}{\sim} N(\mu, \sigma^2) \cdots\cdots (2)$
  (iii) within study variances, s_k^2 are known
  (usually estimated values are plugged in)
- Likelihood based inference combines (1) and (2) to form the marginal model: $\hat{\theta}_k \overset{ind}{\sim} N(\mu, s_k^2 + \sigma^2)$
- Maximum Likelihood Estimate (Frequentist):

$$(\hat{\mu}, \hat{\sigma}) = \arg\max \prod_{k=1}^{K} \left\{ \phi\left((\hat{\theta}_k - \mu)/\sqrt{\sigma^2 + s_k^2} \right) / \sqrt{\sigma^2 + s_k^2} \right\}$$

  where $\phi(\cdot)$ denotes the probability density function of $N(0,1)$
- Monte Carlo (MC) based Posterior Estimate (Bayesian): $\mu|\hat{\theta}_k' s, s_k' s, \sigma^2 \sim N(?, ?)$ and $\sigma^2|\hat{\theta}_k' s, s_k' s \sim$ ARS

Example 1: Does Magnesium sulphate have protective effect after acute myocardial infraction (AMI), particularly through preventing serious arrhythmias?

Analysis Using Estimated log odd-ratios:
log(OR): $\hat{\theta} = (-0.65, -1.02, -1.12, -0.04, 0.21, -2.05, -1.03, -0.30)$
and s.e.(log-OR): $s = (1.06, 0.41, 0.74, 1.17, 0.48, 0.90, 1.02, 0.15)$

R codes:

```
negloglik=function(mu=0,sigma=1){
-sum(dnorm(theta.hat,mean=mu,sd=sqrt(sigma^2+s^2),log=T))}
require(stats4)
fit=mle(negloglik)
#fit=mle(negloglik,method="L-BFGS-B",lower=c(-Inf,1e-08))
summary(fit)
vcov(fit)
confint(fit)
```



```
Results:
        Estimate  Std.Error  2.5%   97.5%
mu    -4.0854e-01  0.1287 -1.1301 -0.1330
sigma -9.8811e-05  0.3102 -0.9482  0.9482
```

Conclusion: Combined estimated log(OR)=-0.4085 with 95% C.I. (-1.1301, -0.1330) and hence magnesium sulphate has protective effect after AMI

- What are the study specific log(OR) estimates?
- What is the predictive distribution of treatment effect in a new trial?
- What is the predictive distribution of the log(OR) to be observed in a new trial?
- We can perform Bayesian meta-analysis to answer these questions

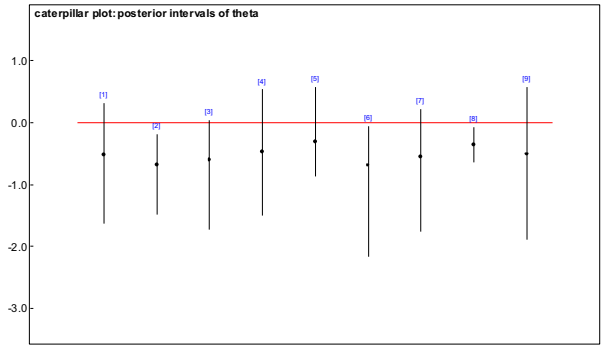
Hierarchical Model: $\hat{\theta}_k \sim N(\theta_k, s_k^2)$ and $\theta_k \sim N(\mu, \sigma^2)$

WinBUGS codes:

```
model{
for(k in 1:K){
theta.hat[k] ~ dnorm(theta[k], tau.s[k])
tau.s[k] <- pow(s[k], -2); theta[k] ~ dnorm(mu, tau)
            }
theta[K+1] ~ dnorm(mu, tau)
tau ~ dunif(0, 1000); mu ~ dnorm(0, 0.0001)
tau <- 1/(sigma*sigma); sigma ~ dexp(1)
    }
Data: http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.1.ISIS-dat.txt
list(theta.hat=c(-0.65,-1.02,-1.12,-0.04,0.21,-2.05,-1.03,-0.30),
s=c(1.06,0.41,0.74,1.17,0.48,0.90,1.02,0.15), K=8)
Inits:
list(mu=0, tau=1)
```

Results:

	mean	sd	2.5%	median	97.5%
mu	-0.5532	0.2833	-1.2160	-0.5250	-0.0466
sigma	0.3971	0.3089	0.0121	0.3352	1.1440
theta[1]	-0.5543	0.4601	-1.6050	-0.5130	0.3340
theta[2]	-0.7211	0.3440	-1.4800	-0.6744	-0.1705
theta[3]	-0.6644	0.4370	-1.7180	-0.5820	0.0649
theta[4]	-0.4783	0.4641	-1.4940	-0.4660	0.5060
theta[5]	-0.2527	0.3551	-0.8530	-0.3044	0.5831
theta[6]	-0.8002	0.5379	-2.1490	-0.6663	-0.0517
theta[7]	-0.6174	0.4782	-1.7530	-0.5471	0.2306
theta[8]	-0.3520	0.1461	-0.6287	-0.3536	-0.0623
theta[9]	-0.5510	0.5734	-1.8220	-0.5026	0.5753



Recall that we have used approximate normality assumption for the estimated log(OR)'s.

Is the assumption reasonable?

- The previous analysis was based on following assumptions:
  - The estimated log(OR) differences are normally distributed
  - The estimated s.e.'s are fixed
- How reasonable these assumptions are when in fact the original observations are binomial counts?
- X_{kj} =# of deaths out of n_{kj} patients in group j of k-th study
- Assume that $X_{kj} \sim Bin(p_{kj}, n_{kj})$ for $j = 1, 2;\ k = 1, \ldots, K$
- Consider a (new!) full Hierarchical model:

$$\begin{aligned} p_{kj} &\sim Beta(m_j\mu_j,\ m_j(1-\mu_j)) \\ \mu_j &\sim Beta(0.5, 0.5), \quad \text{for } j = 1, 2 \\ m_j &\sim Gamma(0.1, 0.1) \end{aligned}$$

- Study-specifc log(OR): $\theta_k = \log \frac{p_{k1}(1-p_{k2})}{p_{k2}(1-p_{k1})}$ for $k = 1, \ldots, K$

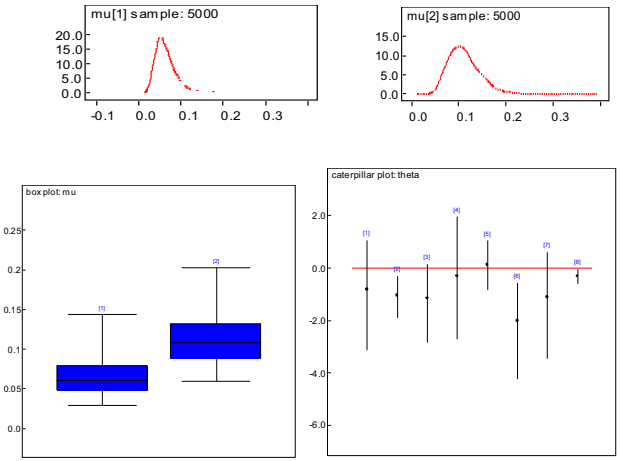
WinBUGS code:

```
model{
for(j in 1:2){
for(k in 1:8){
X[k,j] ~ dbin(p[k,j], n[k,j])
p[k,j] ~ dbeta(a[j], b[j])
          }
a[j] <- m[j]*mu[j]
b[j] <- m[j]*(1-mu[j])
mu[j] ~ dbeta(0.5, 0.5)
m[j] ~ dgamma(0.1, 0.1)
        }
for(k in 1:8){
theta[k] <- logit(p[k,1]) - logit(p[k,2])
          }
log.or <- logit(mu[1]) - logit(mu[2])
    }
Data: http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.1.ISIS-dat.txt
 X[,1] X[,2] n[,1] n[,2]
   1    2    40   36
   9   23   135  135
   2    7   200  200
   1    1    48   46
  10    8   150  148
   1    9    59   56
   1    3    25   23
  90  118  1159 1157
END
Inits:
list(mu=c(0.5, 0.5), m=c(1,1))
```

	mean	sd	2.5%	median	97.5%
log.or	-0.6162	0.5618	-1.6850	-0.6333	0.5582
mu[1]	0.0669	0.0308	0.0294	0.0601	0.1471
mu[2]	0.1135	0.0371	0.0592	0.1083	0.1992
theta[1]	-0.8708	1.0550	-3.1580	-0.8127	1.0260
theta[2]	-1.0530	0.4069	-1.8710	-1.0440	-0.2732
theta[3]	-1.1960	0.7507	-2.7820	-1.1540	0.1821
theta[4]	-0.3417	1.1480	-2.7030	-0.3139	1.9410
theta[5]	0.1195	0.4688	-0.8088	0.1171	1.0420
theta[6]	-2.1150	0.9364	-4.1880	-2.0200	-0.5319
theta[7]	-1.1760	1.0370	-3.4090	-1.1160	0.6349
theta[8]	-0.3067	0.1461	-0.5977	-0.3064	-0.0202

Notice that now the evidence in support of the treatment is no longer as conclusive as before (when estimated log(OR)'s was used)

Moral: *Try to model as many sources of variations as available*



Some Remarks:

- Often an alternative model is used:
  $logit(p_{k1}) = \phi_k + \theta$ and $logit(p_{k2}) = \phi_k$
- If ϕ_k and θ are assumed independent, it forces the treatment risks to be greater than that of control risks
  (notice that $Var(\phi_k + \theta) > Var(\phi_k)$)
- If ϕ_k's are given (fixed) uniform priors, it might lead to inappropriate shrinkage
- Induces strong correlation between treatment (p_{k1}'s) and control (p_{k2}'s) groups when ϕ_k are given vague priors
  (notice that $Corr(\phi_k + \theta, \phi_k) = [1 + Var(\theta)/Var(\phi_k)]^{-1/2}$)
- See Spiegelhalter, Abrams & Myles (2004, p.275) for a prior sensitivity study:
  `http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.2.efm.txt`

Example 2: Is electronic foetal heart rate monitoring clinically effective in reducing the risk of perinatal death?

Traditional normal approximation based model may not be appropriate due to rarity pf perinatal deaths

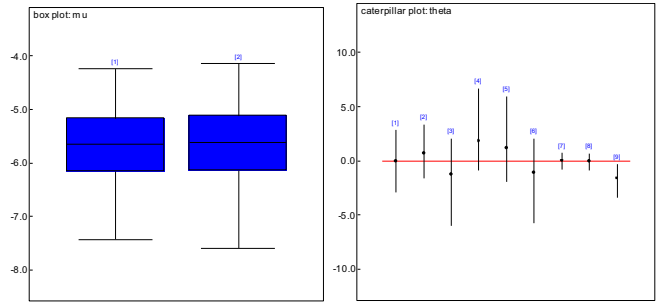
- X_{kj} = # perinatal deaths out of n_{kj} in group j of k-th study
- Assume $X_{kj} \sim Bin(p_{kj}, n_{kj})$ for $k = 1, \ldots, K(= 9),\ j = 1, 2$
- We can use the earlier Beta distribution based model for p_{kj}'s
- However, consider an alternative Hierarchical model:

$$\begin{aligned} logit(p_{kj}) &\sim N(\mu_j,\ \sigma_j^2) \\ \mu_j &\sim N(0, 0.0001), \quad \text{for } j = 1, 2 \\ \sigma_j^2 &\sim InvGamma(0.1, 0.1) \end{aligned}$$

- Study-specifc log(OR): $\theta_k = \log \frac{p_{k1}(1-p_{k2})}{p_{k2}(1-p_{k1})}$ for $k = 1, \ldots, K$

WinBUGS code:

```
model{
for(j in 1:2){
for(k in 1:9){
X[k,j] ~ dbin(p[k,j], n[k,j])
logit(p[k,j]) <- logit.p[k,j]
logit.p[k,j] ~ dnorm(mu[j], tau[j])}
mu[j] ~ dnorm(0, 0.0001)
tau[j] ~ dgamma(0.1, 0.1)
sigma[j] <- 1/sqrt(tau[j])}
for(k in 1:9){theta[k] <- logit.p[k,1]-logit.p[k,2]}
log.or <- mu[1] - mu[2]
sigma.pop <- sqrt(sigma[1]*sigma[1]+sigma[2]*sigma[2])
}

Data: http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.2.efm-dat.txt
n[,1] X[,1] n[,2] X[,2]
 175    1   175     1
 242    2   241     1
 253    0   251     1
 463    3   232     0
 445    1   482     0
 485    0   493     1
6530   14  6554    14
 122   17   124    18
 746    2   682     9
END

Inits:
list(mu=c(0, 0), tau=c(1,1))
```

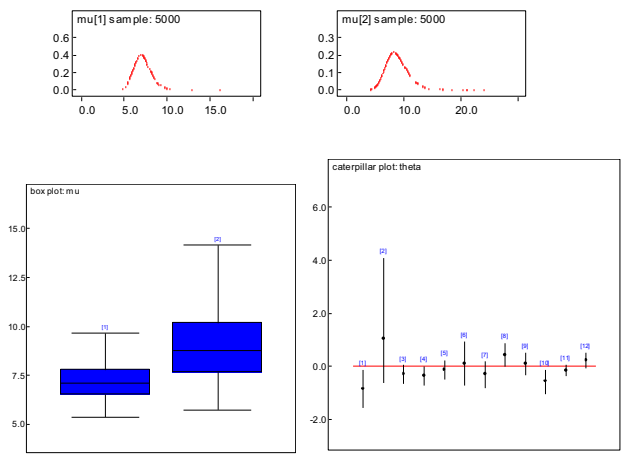
Example 3: Does drug treatment reduce mortality in mild to moderate hypertension adjusting for baseline rates?

- $X_{kj} = \#$ deaths in group j for the k-th study
- $n_{kj} =$ patient-years of follow-up in group j for k-th study
- Consider again a full hierarchical model:

$$\begin{aligned} X_{kj} &\sim Poisson(\lambda_{kj})\ k = 1, \ldots, K, j = 1, 2 \\ \lambda_{kj} &= n_{kj} * \delta_{kj}/1000 \\ \delta_{kj} &\sim Gamma(\mu_j^2 \tau_j, \mu_j \tau_j) \\ \mu_j &\sim Gamma(a, b) \\ \tau_j &\sim Gamma(a, b) \end{aligned}$$

- We would be interested in log of the relative differences: study specific: $\theta_k = \log(\delta_{k1}) - \log(\delta_{k2})$ and population level: $\log(\mu_1/\mu_2)$


```
model{
for(j in 1:2){for(k in 1:12){
X[k,j] ~ dpois(lambda[k,j])
lambda[k,j] <- n[k,j]*delta[k,j]/1000
delta[k,j] ~ dgamma(a[j], b[j])}
a[j] <- b[j]*mu[j]
b[j] <- mu[j]*tau[j]
mu[j] ~ dgamma(0.1, 0.1)
tau[j] ~ dgamma(0.1, 0.1)}
for(k in 1:12){theta[k] <- log(delta[k,1]) - log(delta[k,2])}
log.diff <- log(mu[1]) - log(mu[2])
}

Data: http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.3.hyper-dat.txt
X[,1]   n[,1] X[,2]   n[,2]
 10    595.2    21   640.2
  2    762.0     0   756.0
 54   5635.0    70  5600.0
 47   5135.0    63  4960.0
 53   3760.0    62  4210.0
 10   2233.0     9  2084.5
 25   7056.1    35  6824.0
 47   8099.0    31  8267.0
 43   5810.0    39  5922.0
 25   5397.0    45  5173.0
157  22162.7   182 22172.5
 92  20885.0    72 20645.0
END

Inits:
list(mu=c(1,1), tau=c(1,1))
```



A General Model for Direct Comparison

- In most cases, we will have some outcome events X_{kj} measured in group j of k-th study of sample size n_{kj}
- Assume $X_{kj} \sim f(x|\eta_{kj},\ n_{kj})$ for $k = 1, \ldots, K,\ j = 1, 2$ where η_{kj} denotes the vector of parameters.
- Assume $\eta_{kj} \sim h(\eta|\mu_j, \tau_j)$ where $h(\cdot|\mu, \tau)$ is a conditional density with location μ and scale τ
- $\theta_k = d(\eta_{k1},\ \eta_{k2})$ where $d(\eta_1, \eta_2)$ is some "signed distance" between η_1 and η_2 such that

  (i) $d(\eta_1, \eta_2) = 0$ iff $\eta_1 = \eta_2$ and

  (ii) $d(\eta_1, \eta_2) = -d(\eta_2, \eta_1)$
- The goal would be to obtain posterior distribution of $d(\mu_1, \mu_2)$ or $E(\theta_k|\mu_j' s, \tau_j' s)$ (which is same for all k)

Meta Regression

- Why study results vary systematically?
- Random effects model generally can not identify factors that may explain the sources of variability
- Regression models can be possibly be used to explore reasons why study results vary if study level covariates are available
- Suppose x_k denote a study level covariate for k-th study

$$\hat{\theta}_k = \theta_k + s_k\epsilon_k \quad \text{and} \quad \theta_k = \mu + \beta x_k + \sigma e_k$$

- Marginally (i.e. integrating the random effects), we get

$$\hat{\theta}_k = \mu + \beta x_k + (s_k\epsilon_k + \sigma e_k)$$

- Both frequentist and Bayesian methods can be used to estimate the overall effect μ (wlog assuming $\sum_k x_k = 0$)

(source: Sutton and Abrams (2001), *Stat Meth in Med Res*, **10**, p.277-303)

Indirect Comparison Studies

- Suppose an established treatment C (active-control) exists for a severe health condition
- Suppose a new treatment T is being evaluated to treat that health condition
- The efficacy of T would ideally be estimated using a randomized control trial (RCT) with a placebo P
- But...existence of C may make the use of placebo unethical
- In this case the efficacy of T may have to estimated *indirectly* using (past) data on comparisons between C and P
- Can we compare a new treatment T and control C without using a RCT that directly compares T with placebo P?

- More generally, can we make comparisons between (several) treatments that may well never have been directly compared?
- Is it really possible to draw inferences on the treatment effects compared with a control only?
- On a positive note see Song et al. (2003) article: `http://www.bmj.com/content/326/7387/472.full`
- On cautionary notes see J. A. Berlin's talk (04/27/2010): `http://www.cceb.upenn.edu/biostat/conferences/ClinTrials10/`
- Suppose η_{kj} represents expected response of treatment j being given in study k (control is labeled as $j = 0$)
- A simple model: $\eta_{kj} = \theta_k + \phi_{kj}$
- Often it is convenient to assume $\phi_{kj} \sim N(\mu_j, \sigma_j^2)$
- A variety of other possible models can be considered

Example 4: Can we compare alternative therapies for lowering blood pressure by estimating effects that have never been directly measured?

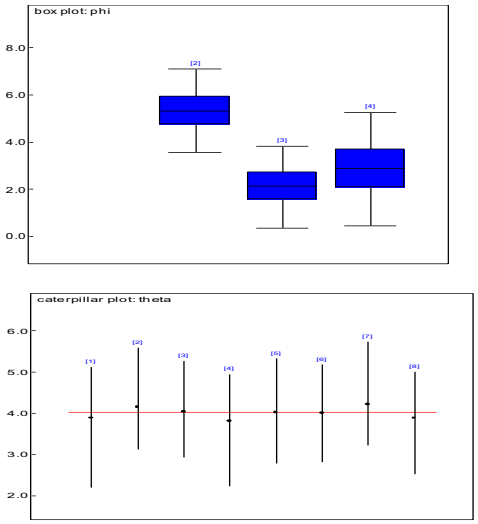
- Let X_{kj} = mean change in blood pressure for the j-th treatment in the k-th study where $j = 0, 1, 2, 3$ and $k = 1, \ldots, K = 8$
- Four treatments (control, A, B, C) have been given in a set of crossover experiments comprising RCTs and single-arm studies
- But...no direct comparison between treatments A and B made
- And...we are interested in this particular contrast (i.e., A vs. B)
- Let $X_{kj} \sim N\left(\eta_{kj}, \frac{\sigma_j^2}{n_{kj}}\right)$ for $k = 1, \ldots, K = 8,\ j = 0, \ldots, J = 3$
- Further assume $\frac{(n_{kj}-1)S_{kj}^2}{\sigma_j^2} \sim \chi^2_{n_{jk}-1}$
- Let $\eta_{jk} = \theta_k + \phi_j$ where $\phi_0 = 0$
- Thus, θ_k = "control" in k-th study and ϕ_1, ϕ_2, ϕ_3 measure mean effects of A, B and C, respectively.

WinBUGS code:

```
model{
for(i in 1:I){
x[i] ~  dnorm(mu[i],prec[i])
mu[i] <- phi[treat[i]] + theta[study[i]]
prec[i] <- n[i]/(sigma[treat[i]]*sigma[treat[i]])
SS[i] <- s[i]*s[i]*(n[i]-1)
SS[i] ~  dgamma(a[i],b[i])
a[i] <- (n[i]-1)/2
b[i] <- 1/(2*sigma[treat[i]]*sigma[treat[i]])
            }
for(k in 1:K){
theta[k] ~ dnorm(mu.theta, inv.sigma2.theta)
             }
#contrasts of interest
   AvB  <- phi[2]-phi[3]

mu.theta ~ dunif(-50,50)
inv.sigma2.theta  <- 1/(sigma.theta*sigma.theta)
sigma.theta ~ dunif(0,100)
for(j in 1:J){
log(sigma[j]) <-logsigma[j]
logsigma[j]~dunif(-10,10)
            }
phi[1]  <- 0
for(j in 2:4){
phi[j] ~ dunif(-50, 50)
            }

}
```



```
Data: http://www.mrc-bsu.cam.ac.uk/bayeseval/ex8.4.blood-dat1.txt
list(I=14, K=8, J=4)

n[]  x[]   s[] treat[] study[]
 41  8.90  7.49  2  1
 39  6.05 10.28  4  1
 47  5.51  8.72  1  2
100  6.21  8.02  3  2
 53  3.75  7.07  1  3
 54 10.20  9.39  2  3
 47  3.04  9.20  1  4
 44  8.43  8.17  2  4
 30  2.97  7.69  1  5
 32  6.53  7.80  3  5
 32  7.78  6.78  4  5
 69  3.99  8.04  1  6
 68  5.28  7.58  1  7
 67  3.34  8.01  1  8
END

Inits:
list(mu.theta=4,sigma.theta=1,logsigma=c(2,2,2,2),phi=c(NA,4,4,4))
```



Multiple Testing

- Consider m tests: H_{0j} vs. H_{1j} for $j = 1, \ldots, m$
- Suppose that the tests are conducted independently each at level α; i..e., the probability of declaring a particular test is significant under its null is α
- But the probability of declaring at least 1 of the m tests significant is $1 - (1 - \alpha)^m$
- E.g., if we use $\alpha = 0.05$, we have

$m =$	5	10	30	50	100
$1 - 0.95^m$	0.23	0.40	0.79	0.92	0.99

- The probability of declaring at least one test significant is almost a certainty if when we have 100 tests (each at level 0.05)

Consider the following set-up:

	# H_0 not rejected	# H_0 rejected	total
# true H_0	U	V	m_0
# false H_0	T	S	m_1
total	$m - R$	R	m

- Unfortunately U, V, T and S are all unobservable; only R (and obviously m) is observable
- Some notions/definitions:
  - Per-comparison error rate: $PCER = \frac{E[V]}{m}$
  - Family-wise error rate: $FWER = \Pr[V \geq 1]$
  - False discovery rate: $FDR = E\left[\frac{V}{\max\{R,1\}}\right]$
  - Positive FDR: $pFDR = E\left[\frac{V}{R} | R > 0\right] = FDR / \Pr[R > 0]$

- The goal is now to control some these error rates
- Most often procedures are based on adjusting the (unadjusted) p-values, p_j's
- Bonferroni: Reject any H_{0j} with p-value $\leq \frac{\alpha}{m}$
  i.e., adjusted p-value $= \min\{mp_j, 1\}$
- Sidak: Reject any H_{0j} with p-value $\leq 1 - (1 - \alpha)^{1/m}$
  i.e. adjusted p-value $= \min\{1 - (1 - p_j)^m, 1\}$
- Bonferroni and Sidak performs very similar; however both are usually too conservative.
- Holm step-down: Order the unadjusted p-values as
  $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
  adjusted p-value $= \max_{1 \leq k \leq j} \min\{(m - k + 1)p_{(k)}, 1\}$
- Holm's approach is less conservative than Bonferroni/Sidak

- All of these adjusted p-values attempt to control FWER
- Westfall and Young step-down approach:
  adjusted p-value $= \max_{1\le k\le j} \Pr[\min_{k\le l\le m} p_{(l)} \le p_{(k)} | H_0^c]$
- Benjamini and Hochberg:
  adjusted p-value $= \min_{j\le k\le m} \min\{\frac{mp_{(k)}}{k}, 1\}$
- Asymptotically, as m becomes large (under independence of tests) it can be shown that

$$FDR \approx pFDR \approx \frac{E[V]}{E[R]}$$

  where the last ratio is the proportion of false discoveries (PFD)
- There are "adaptive" modifications of Benjamini and Hochberg procedure:
  Compute $\hat{m} = \max\{i : p_{(i)} \le \frac{\alpha}{\hat{p}_0}\frac{i}{m}\}$ and reject H_{0j} if $p_{(j)} \le p_{(\hat{m})}$

Some online resources:

- HHS:
  http://www.hhs.gov/recovery/programs/cer/execsummary.html
- Spiegelhalter, D. J. (2004):
  http://projecteuclid.org/euclid.ss/1089808280
- Spiegelhalter, Abrams and Myles (2003) Book (Chap.8):
  http://www.mrc-bsu.cam.ac.uk/bayeseval/
  [*All four examples in this talk are adapted from the above book*]
- Dmitrienko, Tamhane and Bretz (2009) *Multiple Testing Problems in Pharmaceutical Statistics*
  http://www.crcpress.com/product/isbn/9781584889847

THANKS!

Dhanyavaad धन्यवाद