# Bayesian Meta-Analysis

*Biostat Working Group*

Chuan Zhou

chuan.zhou@vanderbilt.edu

Department of Biostatistics
Vanderbilt University

# Bayes' theorem

- Let $\theta$ be a parameter of interest and $y$ denote the observed data.
- Let $\pi(\theta)$ denote the *prior distribution* and $\pi(y)$ denote the *likelihood*.
- Bayes' theorem:
  - posterior distribution given by

$$\pi(\theta \mid y) = \frac{\pi(\theta)\pi(y \mid \theta)}{\int \pi(\theta)\pi(y \mid \theta) d\theta} \propto \pi(\theta)\pi(y \mid \theta)$$

  - application of the law of probability
  - denominator ensures the posterior is a proper distribution
  - denominator is also called normalizing constant, contains no information about $\theta$

# Bayesian statistical paradigm

- Use of 'subjective' or 'degree of belief' interpretation of probability
- Bayesian statistical analysis has three main components:
  - *prior distribution*
  - *likelihood (sampling distribution)*
  - *posterior distribution*
- Initial beliefs, which are quantified using the prior distribution, are combined with information in the data, quantified by the likelihood, to produce new/updated beliefs in the form of the posterior distribution.
- Duality between sample and distribution; generate sample from posterior distribution

# BUGS language

- Bayesian inference Using Gibbs Sampling
- Intended for problems with no analytic solution, or difficult to approximate
- Describe the joint probability model (graphical model)
- Workhorse under the hood: Markov chain Monte Carlo (MCMC), Gibbs sampler, Metropolis-Hastings' sampler, Slice sampler
- Generate dependent samples sequentially, and eventually the samples would come from the target distribution
- Syntax similar to S/R language
- JAGS - just another Gibbs sampler; same functionality of classic BUGS

# *Meta-analysis*

- A *meta-analysis* is a statistical analysis of a collection of studies:
  - studies are the primary units of analysis
  - focus on contrasting and combining results from different studies
- Part of a broader goal of *research synthesis*
  - bring together current knowledge on a given topic
  - narrative components
  - quantitative components

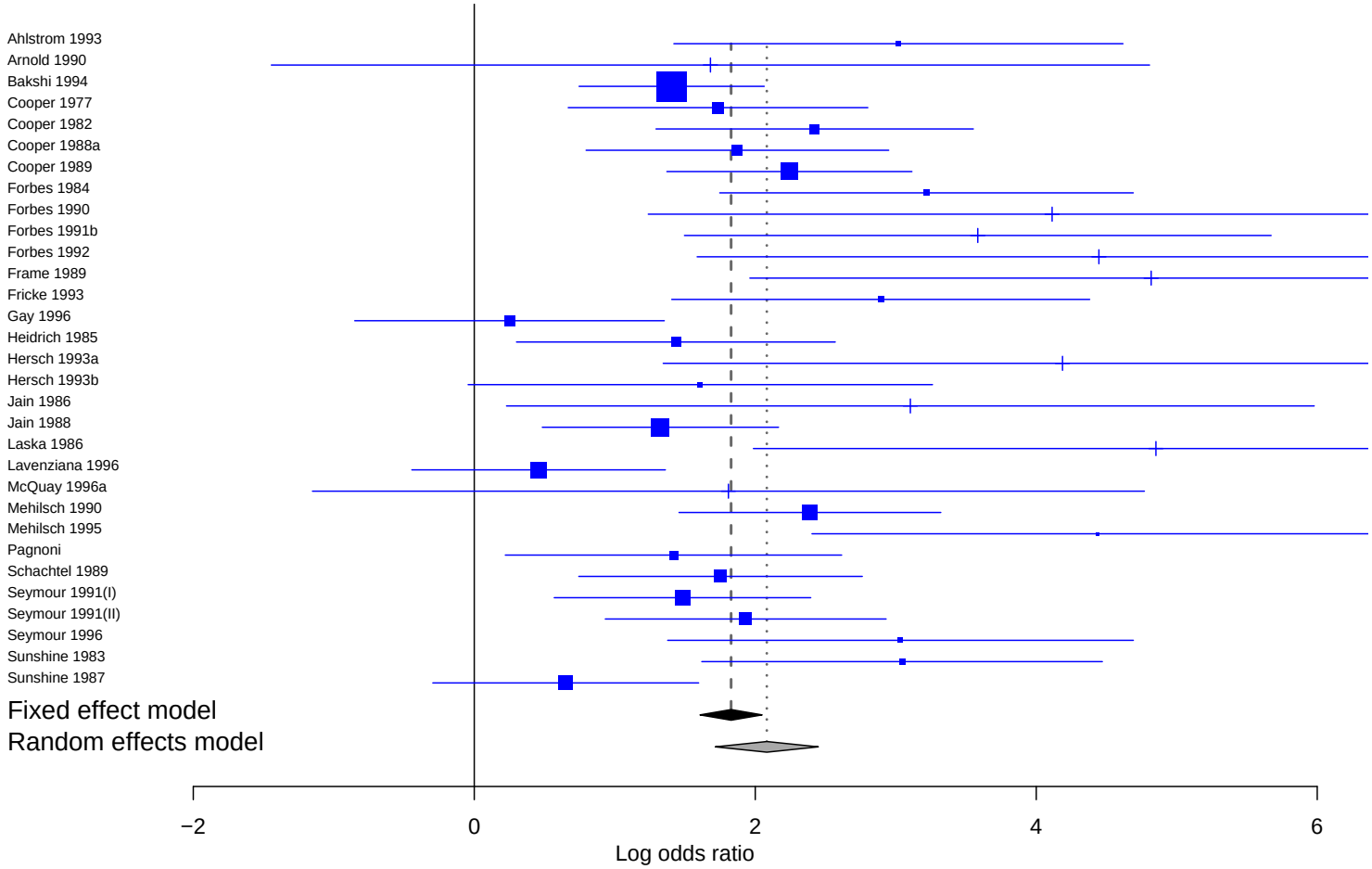
# Meta-analysis cont'd

- Ideal:
  - access to **all** studies conducted on specific topic
  - subject-specific information/data
- Primary difficulties
  - heavy reliance on published literature
  - publication bias; authors as well as editors
  - quality/extent of information in published articles

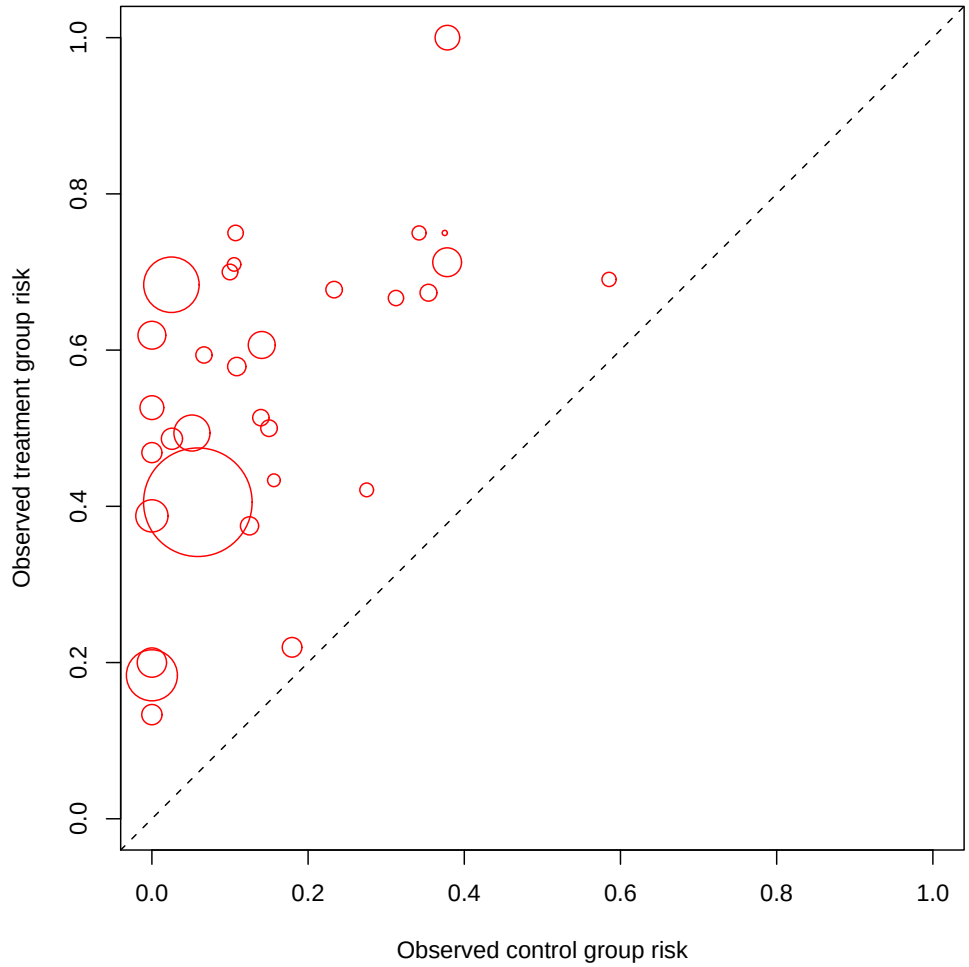
# Example

- Single-dose ibuprofen for post-operative pain
  - non-steroidal anti-inflammatory (NSAID) analgesic
  - 46 placebo-controlled trials between 1997 and 1996
- Warn, Thompson and Spiegelhalter (2002)
- Outcome: at least 50% pain relief in 4-6 hours after administration
- Exposure: doses range from 50mg to 800mg
- Initial impression
  - considerable evidence that ibuprofen improves pain relief
  - heterogeneity in effect sizes and uncertainty
  - sample sizes vary from 28 to 391; median 80

# Forest plot


Ahlstrom 1993
Arnold 1990
Bakshi 1994
Cooper 1977
Cooper 1982
Cooper 1988a
Cooper 1989
Forbes 1984
Forbes 1990
Forbes 1991b
Forbes 1992
Frame 1989
Fricke 1993
Gay 1996
Heidrich 1985
Hersch 1993a
Hersch 1993b
Jain 1986
Jain 1988
Laska 1986
Lavenziana 1996
McQuay 1996a
Mehilsch 1990
Mehilsch 1995
Pagnoni
Schachtel 1989
Seymour 1991(I)
Seymour 1991(II)
Seymour 1996
Sunshine 1983
Sunshine 1987
Fixed effect model
Random effects model
−2
0
2
4
6
Log odds ratio

# L'Abbe plot


Observed treatment group risk
0.0
0.2
0.4
0.6
0.8
1.0
Observed control group risk
0.0
0.2
0.4
0.6
0.8
1.0

# Fixed-effect model on summary statistics

Consider log odds ratio (LOR), $d_i = \log(\frac{r_i^t(n_i^c - r_i^c)}{r_i^c(n_i^t - r_i^t)})$ is the observed log odds ratio for study $i$, $s_i^2 = \text{var}(d_i) = \frac{1}{r_i^t} + \frac{1}{n_i^t - r_i^t} + \frac{1}{r_i^c} + \frac{1}{n_i^c - r_i^c}$ is the *known* variance of $d_i$. For moderately large study sizes, we can assume for $i = 1, \cdots, k$,

$$d_i \sim_{indep.} \text{N}(\theta, s_i^2).$$

- data: $\boldsymbol{y} = (d_1, s_1^2, \cdots, d_k, s_k^2)$
- parameter of interest: $\theta$ = average treatment effect
- likelihood: normal density $f(y_i \mid \theta)$
- target: posterior distribution $\pi(\theta \mid \boldsymbol{y})$
- need to specify prior on $\theta$

# The code

```
model {
   for (i in 1:k) {
      d[i] ~ dnorm(theta, precision.d[i]); # likelihood
      precision.d[i] <- 1/variance.d[i];
   }
   theta ~ dnorm(0, 1e-3);  # prior
}
```

# Random-effects model

Assume study summary statistic $d_i$ is a draw from a distribution with a study-specific mean $\theta_i$, and variance $s_i^2$:

$$d_i \,|\, \theta_i, s_i^2 \sim_{indep.} \mathrm{N}(\delta_i, s_i^2),$$

$$\delta_i \,|\, \theta, \tau^2 \sim_{indep.} \mathrm{N}(\theta, \tau^2).$$

- parameters $\boldsymbol{\theta} = (\delta_1, \cdots, \delta_k, \theta, \tau^2)$: $\theta$ is population average, $\delta_i$ is study-specific effect, $\tau^2$ is between study variation.
- $p(\boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto p(\boldsymbol{y} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta}) = \{\prod_{i=1}^{k} p(d_i \,|\, \delta_i, s_i^2) p(\delta_i \,|\, \theta, \tau^2)\} p(\theta, \tau^2)$
- need to specify priors on $\theta$ and $\tau^2$

# The code

```
model {
   for (i in 1:k) {
      d[i] ~ dnorm(delta[i], precision.d[i]);
      precision.d[i] <- 1/variance.d[i];
      delta[i] ~ dnorm(theta, precision.tau);
   }
   precision.tau ~ dgamma(0.001, 0.001);
   tau <- sqrt(1/precision.tau);
   theta ~ dnorm(0, 0.01);
}
```

# Models on odds ratio scale

What if the study sizes are small and normal assumption does not hold? Try the following

$$
\begin{aligned}
r_i^c &\sim \text{Bin}(n_i^c, \pi_i^c) \\
r_i^t &\sim \text{Bin}(n_i^t, \pi_i^t) \\
\mu_i &= \text{logit}(\pi_i^c) \\
\text{logit}(\pi_i^t) &= \mu_i + \delta_i \\
\delta_i &\sim \text{N}(\theta, \tau^2)
\end{aligned}
$$

in which $\delta_i = \text{logit}(\pi_i^t) - \text{logit}(\pi_i^c)$ is the log-odds ratio. Need priors on $\pi_i^c$, $\theta$ and $\tau^2$.

# Code for method (a)

```
model {
      for (i in 1:k) {
          r.c[i] ~ dbin(pi.c[i], n.c[i]);
          r.t[i] ~ dbin(pi.t[i], n.t[i]);
          mu[i] <- logit(pi.c[i]);
          logit(pi.t[i]) <- mu[i] + delta[i];
          delta[i] ~ dnorm(theta, precision.tau);
          # pi.c are unrelated, fixed effect
          pi.c[i] ~ dunif(0, 1);
      }
      theta ~ dnorm(0, 0.1);
      precision.tau <- 1/(tau*tau);
      tau ~ dunif(0, 2);
}
```

# Code for method (b)

```
model {
      for (i in 1:k) {
          r.c[i] ~ dbin(pi.c[i], n.c[i]);
          r.t[i] ~ dbin(pi.t[i], n.t[i]);
          mu[i] <- logit(pi.c[i]);
          logit(pi.t[i]) <- mu[i] + delta[i];
          delta[i] ~ dnorm(theta, precision.tau);
          # pi.c follow common distribution therefore similar
          pi.c[i] ~ dbeta(alpha, beta);
      }
      theta ~ dnorm(0, 0.01);
      precision.tau <- 1/(tau*tau);
      tau ~ dunif(0, 2);
      alpha ~ dunif(1, 100);
      beta ~ dunif(1, 100);
}
```

# Adjusting for covariates

Q: How did the dose level $x_i$ of ibuprofen influence the size of the treatment effect?

On the log-odds ratio scale, we can investigate this by fitting a model in which

$$\delta_i = \delta_i^* + \gamma(\log(x_i) - \frac{1}{k}\sum_{i=1}^{k}\log(x_i))$$

where $\delta_i^* \sim \mathrm{N}(\theta, \tau^2)$. Here $\tau^2$ measures the amount of residual heterogeneity not explained by the covariate.

# The code for method (a)

```
model {
      for (i in 1:k) {
          r.c[i] ~ dbin(pi.c[i], n.c[i]);
          r.t[i] ~ dbin(pi.t[i], n.t[i]);
          mu[i] <- logit(pi.c[i]);
          logx[i] <- log(x[i]);
          logit(pi.t[i]) <- mu[i] + delta[i];
          delta[i] <- delta.star[i]+gamma*(logx[i]-mean(logx[]));
          delta.star[i] ~ dnorm(theta, precision.tau);
          pi.c[i] ~ dunif(0, 1); # pi.c are unrelated, fixed effect
      }
      theta ~ dnorm(0, 0.1);
      precision.tau <- 1/(tau*tau);
      tau ~ dunif(0, 2);
      gamma ~ dnorm(0, 1e-4);
}
```

# A script file for running BUGS

```
model in OR-xa.bug
data in meta-x.dat
compile
parameters in meta-x.in
initialize
update 10000
monitor set theta, thin(5)
monitor set tau, thin(5)
monitor set gamma, thin(5)
update 10000
coda *, stem(ibu)
exit
```

# Summary

- Flexible models for meta-analysis
- Familiarize with BUGS language and Bayesian inference
- Focus on posterior distribution
- Much is not covered, in particular MCMC, Bayesian model selection, convergence diagnostic, etc.
- Useful R packages for meta-analysis: rmeta and meta
- MCMCpack R package can handle Bayesian generalized linear model nicely
- JAGS is a nice alternative to openBUGS under linux
- Use coda or BOA R packages for posterior processing