Bayesian Methods and Meta-Analysis

John W. Seaman, Jr., Ph.D.
Professor of Statistics

Department of Statistical Science
Baylor University
Waco, Texas

October 2011

Abstract

Bayesian methods provide for a synthesis of evidence admitting multiple sources of information and a full accounting of model uncertainty. In the last 15 years, advances in computation have allowed the development of increasingly complicated, fully Bayesian approaches to a wide spectrum of biostatistical problems. Bayesian methods for the meta-analytic synthesis of evidence have received considerable attention in the biomedical literature. In this presentation, we briefly introduce Bayesian methods and summarize their advantages in meta analysis.

Outline

- Overview of Bayesian Inference
- Example: Bayesian Analysis of NRS Probability
- Bayesian Meta-Analysis
- Example: Meta-Analysis of NRS Studies
- Advantages and Disadvantages
- Appendix on Notation
- References

Motivation for Bayesian Methods

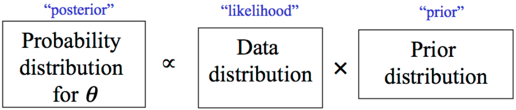
- Bayesian inference is becoming increasingly prevalent in medical/biopharmaceutical literature
- Increasing interest in Pharma and FDA
- Bayesian statistical methods provide a formal method for learning from experience. They have many advantages:
  - Results are easily interpreted in terms of probabilities
  - ALL of the relevant data, past and present, can be utilized
  - New results "update" past results: "Today's posterior becomes tomorrows prior"
  - Straight-forward to handle complex models

Combining New and Old Information: What We Want

- What we want:
  - To answer questions about the effect size or other *parameters* of interest (shorthand: $\boldsymbol{\theta}$), and
  - Account for our uncertainty in those answers.
- Information our answers should be able to make use of:
  - What we know from the current study (or studies), and
  - What we knew before the study (or studies), including
    - Prior data (before the current study (or studies))
    - Expert opinion
- How do we combine the study data with prior information?

Combining New and Old Information: Bayes' Theorem

- Mathematically, data from the current study (or studies) is represented by the *likelihood function* or *data model* and what we knew before by the *prior probability distribution*.
- We combine study data (the likelihood) and prior information (the prior) with a result called Bayes' theorem:



- The result (the left-hand-side above) is called the *posterior distribution*–it is what we can say *after* we make use of prior information and study data.

Non-Resorative Sleep Example

- Patients are given a benzodiazepine-like drug as a sleeping aid.
  - A complication is that some patients experience non-restorative sleep (NRS) during treatment.
  - This type of complication is well-studied for some populations, but not your sub-population of interest, say, hypertensive males aged 50-60.
  - You suspect that patients in this sub-population are more likely than not to experience NRS under this treatment.
- A small exploratory study of 12 similar hypertensive males, aged 50-60, were treated with the drug.
  - Of these, 9 experienced NRS.
  - This seems consistent with your suspicion that these patients are vulnerable.
  - Is it statistically significant?

NRS Example: Non-Bayesian Approach

- Data: $n = 12$ patients; Number w/ NRS, $X = 9$;
- Assumption: $P(\text{NRS}) = \theta$, unknown & constant;
- Subjects are independent;
- Want to test $H_0 : \theta = 0.5$ *vs.* $H_1 : \theta > 0.5$;
- Data model: Use the *binomial data model*:

$$\text{P-value: } P(X \geq 9 | \theta = 0.5) = 0.075.$$

- Interpretation: **IF** we sampled from this same population ($\theta = 0.5$) repeatedly, each time with $n = 12$, then the chance we would see $X = 9$ or anything more extreme is 0.075. So, not significant.[1]

[1] Technical note: If this had been a negative binomial experiment, the P-value would have been 0.0325!

Bayes Approach: Prior Info About NRS Prob, θ

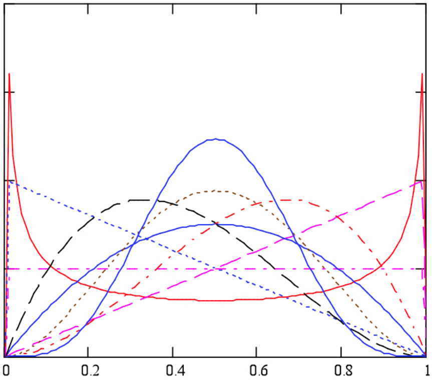


Figure: Choices for prior distributions on the NRS probability, θ. These are all from the beta family of distributions. We must choose one of these using prior data, expert opinion, or both. Note how we treat θ as a random variable. (The "flat" one is also called a uniform distribution.)

Bayesian Approach: Possible Origins of Study Data

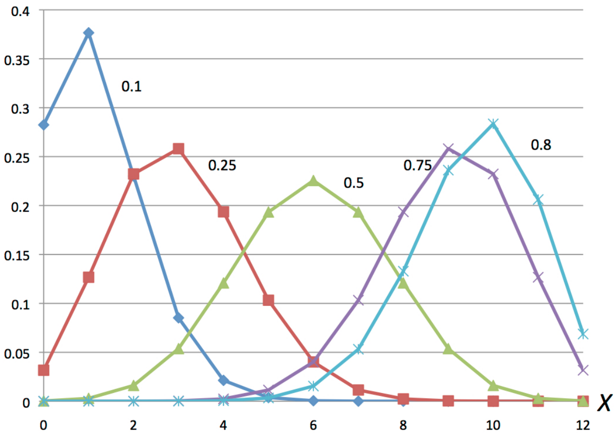


Figure: Different binomial data models. Labels are values of male NRS probability, θ. We don't know which one produced our data!

Bayesian Approach: Likelihood

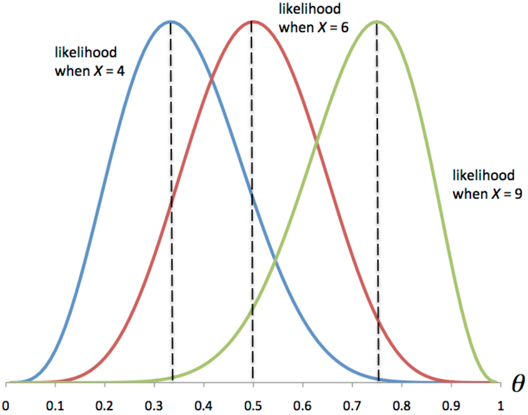
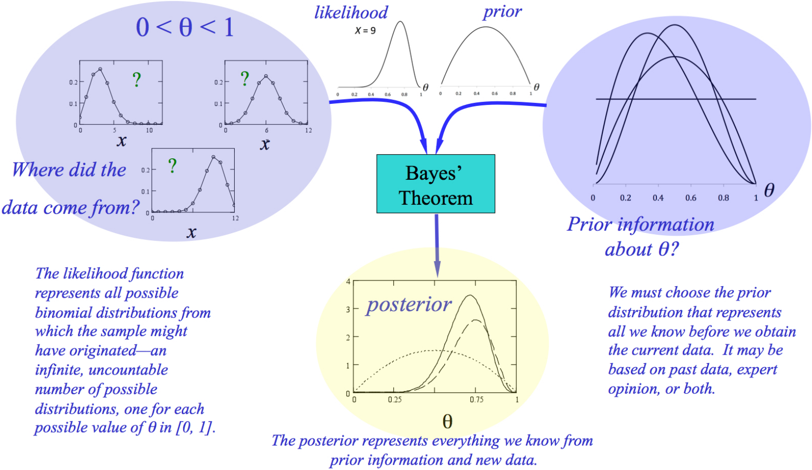
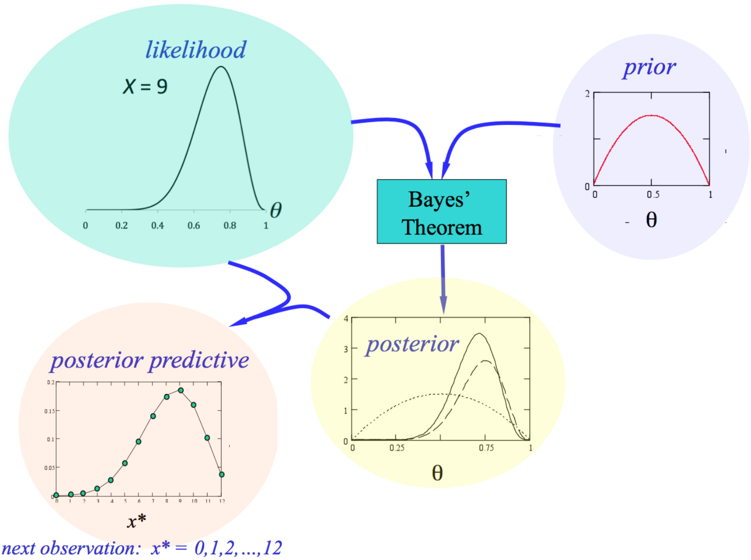


Figure: Binomial likelihood functions for different numbers of NRS males. Labels are values of male NRS count, X. We observed $X = 9$, so the most likely value of θ is 0.75.

Bayesian Approach: General Structure for Inference

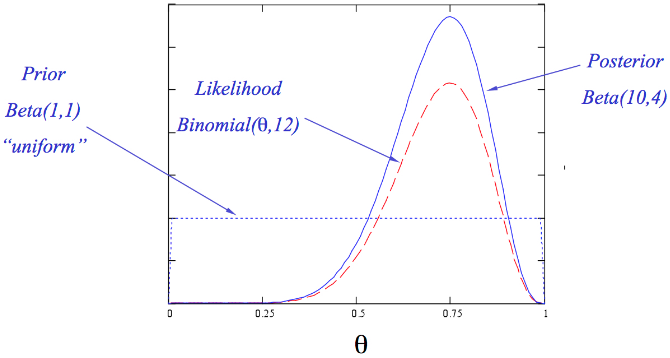


Bayesian Approach: General Structure for Prediction



NRS Example: Bayesian Analysis

- Same experimental set-up as before.[2]
- Use a uniform ("flat") prior on θ.
- Bayes' theorem yields the results below:



[2] Technical note: Does not matter if the data model is binomial or negative-binomial!

NRS Example: Bayesian Analysis w/ Different Priors

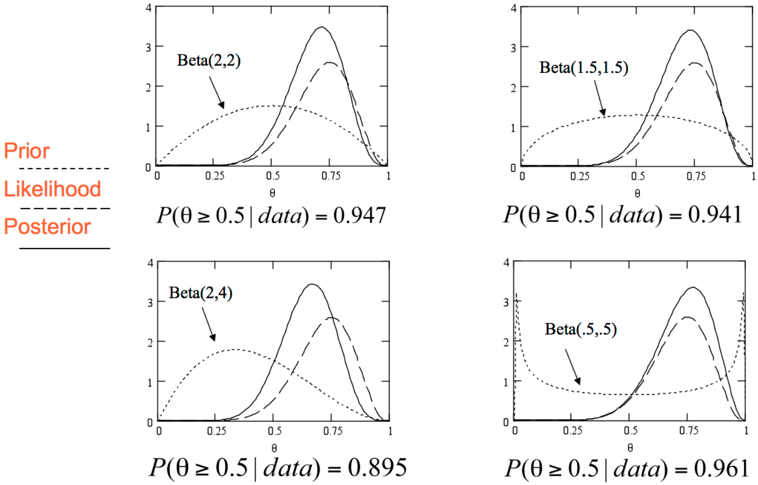
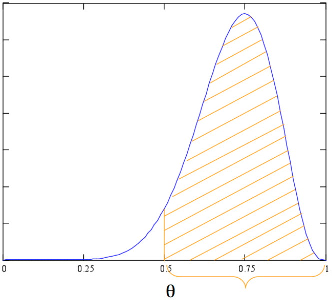


Figure: Using 4 different priors–very similar results with only $n = 12$.

NRS Example: Bayesian Results ("flat" prior again)

- What is the probability that NRS probability exceeds 50%?

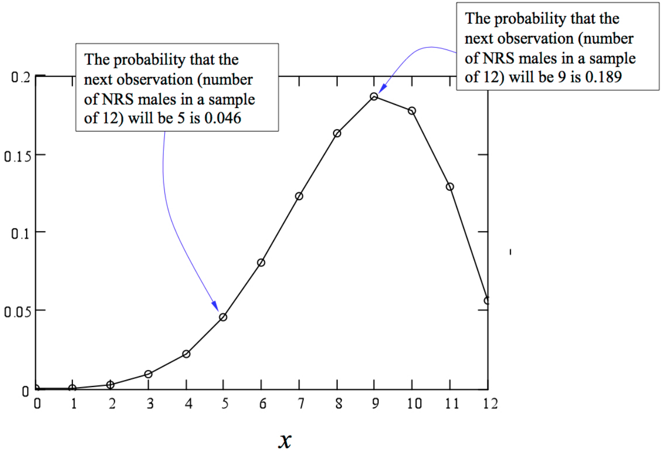
$$P(\theta \geq 0.5 | X = 9) \approx 0.954$$



- This is the posterior probability of the alternative hypothesis. It does *not* depend on hypothetical repeated sampling, but conditions on the data in hand.

NRS Example: Posterior Predictive Distribution

- We can ask questions about future observations...



Using the Posterior as the Next Prior

- A huge advantage of the Bayesian approach its sequential "learning" nature.
- The posterior we just obtained on the probability of NRS for this population can now be used as the prior for that probability in a new, larger study.
- This is one motivation for starting with a "flat" prior, as we did above.

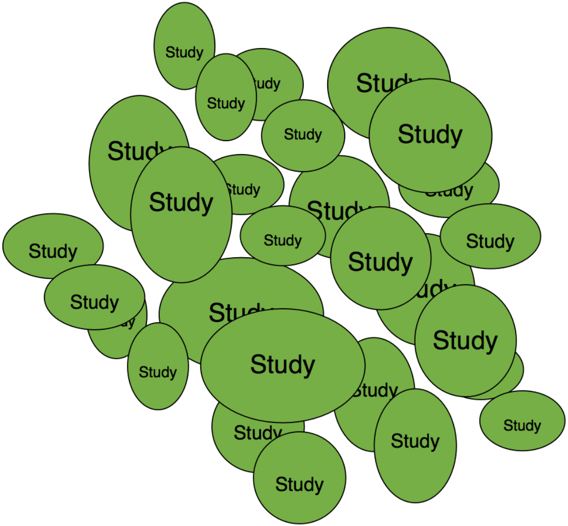
Bayesian Science: Today's Posterior $\rightarrow$ Tomorrow's Prior

- Ideal Bayesian inference cycle:
  - Prior data and/or expert opinion $\rightarrow$ prior distribution, $p(\boldsymbol{\theta})$
  - Combine $p(\boldsymbol{\theta})$ with new data, $\mathbf{x} \rightarrow$ posterior, $p(\boldsymbol{\theta}|\mathbf{x})$
  - New questions...new hypotheses...plan new experiments...
  - Old posterior, $p(\boldsymbol{\theta}|\mathbf{x})$, becomes the new prior, call it $p^*(\boldsymbol{\theta})$
  - Combine $p^*(\boldsymbol{\theta})$ with new data, $\mathbf{x}^* \rightarrow$ new posterior, $p^*(\boldsymbol{\theta}|\mathbf{x}^*)$
  - And so on: prior $\rightarrow$ posterior $\rightarrow$ prior $\rightarrow$ posterior...
- Problems (all solveable):
  - $\boldsymbol{\theta}$ may change;
  - data model for $\mathbf{x}$ (the likelihood) may change;
  - posterior "too precise" to become new prior.

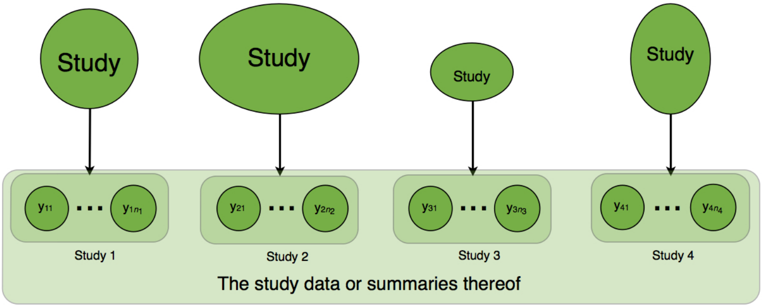
Cost of Using Bayesian Methods

- Model parameters (eg., effect size, variance components) must have prior distributions. These can be difficult and expensive to obtain.
- Advances in computational abilities have allowed increasingly flexible Bayesian modeling, but this requires technical expertise and specialized software (eg. SAS Proc MCMC, WinBUGS, R2WinBUGS).
- Interpretation different from “traditional” statistical methods.
    - Interval estimates and test of hypotheses have different interpretation than confidence intervals and P-values.
    - Bayes results are in terms of probability statements. For example “the probability that mean change from baseline at week 12 is larger than zero is 0.93”.

Meta Analysis: Multiple Candidate Studies (what we see)



Selected Studies (what we see)



There will be variability in study outcomes. How do we model this?

Example: Relative NRS Risk for BL Users vs. Non-Users

- Studies investigate the relative risk of "experiencing NRS" for benzodiazepine-like (BL) drug users vs. non-users.
- Each is a RCT with similar definition of the event.
- Each study reports a 2×2 table of "successes and failures" for both a treatment and a control group.
- You have access to summary results in the form of sample sizes, numbers of subjects experiencing NRS, odds ratios, etc.
- Suppose 13 studies pass your inclusion criteria...

For each NRS study ($s = 1, 2, \ldots, 13$)

- Treatment: NRS rate π_1^s (unknown)
  - sample size, n_1^s
  - x_1^s, the number with NRS out of n_1^s
- Control: Baseline rate, π_0^s (unknown)
  - sample size, n_0^s
  - x_0^s, the number with NRS out of n_0^s
- We would like to know the relative risk:

$$RR = \pi_0^s / \pi_1^s$$

for each study and overall.

Summary of 13 Selected Studies

Study Summary Data[3]

		Con			Treat		
Study	NRS	Days	$O_1 \equiv$	NRS	Days	$O_0 \equiv$	Odds Ratios
s	x_0^s	n_0^s	$\frac{x_0^s}{n_0^s - x_0^s}$	x_1^s	n_1^s	$\frac{x_1^s}{n_1^s - x_1^s}$	O_0/O_1
1	55	97	1.310	67	101	1.971	0.665
2	137	282	0.945	187	306	1.571	0.601
3	505	927	1.197	590	915	1.815	0.659
4	62	123	1.016	74	140	1.121	0.907
5	118	236	1.000	99	239	0.707	0.707
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

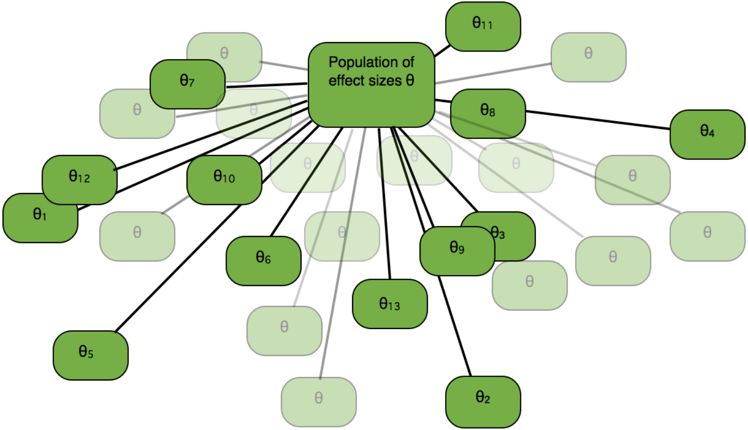
[3] Modified from Parmigiani (2002), p. 128 ff. My modification is hypothetical. I have switched control & treatment from the original and used 13 of his 14 studies in this modification.

Outcome Heterogeneity

		Con			Treat		
Study	NRS	Days	$O_1 \equiv$	NRS	Days	$O_0 \equiv$	Odds Ratios
s	x_0^s	n_0^s	$\frac{x_0^s}{n_0^s - x_0^s}$	x_1^s	n_1^s	$\frac{x_1^s}{n_1^s - x_1^s}$	O_0/O_1
1	55	97	1.310	67	101	1.971	0.665
2	137	282	.945	187	306	1.571	0.601
3	505	927	1.197	590	915	1.815	0.659
4	62	123	1.016	74	140	1.121	0.907
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

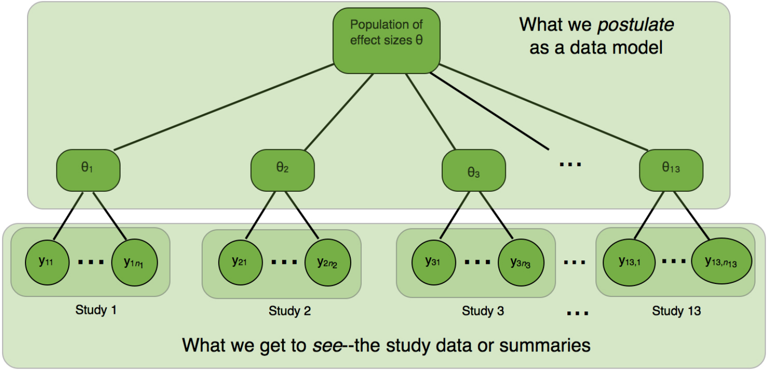
- There is variability in the odds of NRS among the trials.
- This is due to different follow-up times, different patient selection criteria, etc.
- Question: How do we model this variation?

Multiple Candidate Studies (what we imagine)

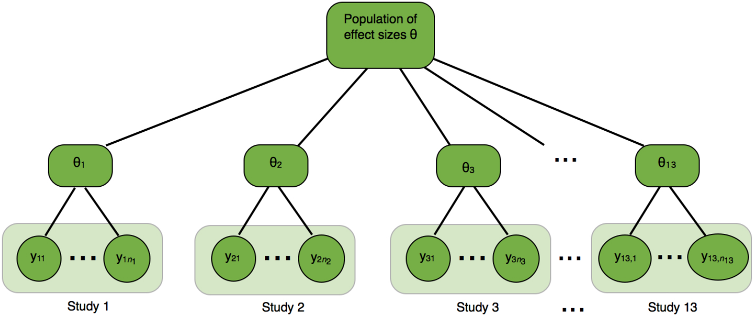


Here the θ's are NRS probabilities: $\theta_s = (\pi_0^s, \pi_1^s)$

Hierarchical Model for 13 Studies (what we see & imagine)



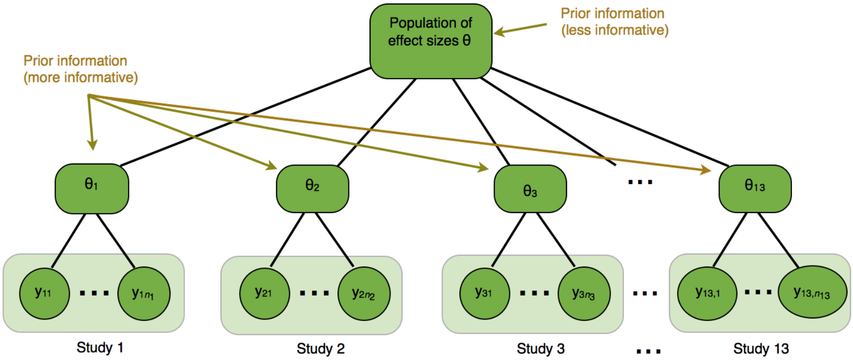
Hierarchical Model for 13 Studies



As we move up the hierarchy, we know less about parameters . . .

Hierarchical Prior Structure

...so, priors are more informative "closer to the data":



For each NRS study ($s = 1, 2, \ldots, 13$)

- Treatment: NRS probability, π_1^s, unknown
  - sample size, n_1^s
  - x_1^s, the number with NRS out of n_1^s
- Control: Baseline NRS probability, π_0^s, unknown.
  - sample size, n_0^s
  - x_0^s, the number with NRS out of n_0^s
- It seems like we should get priors on π_0^s and π_1^s. Instead, for technical reasons, we consider the log odds ratios:

$$\lambda^s = \log\left(\frac{\pi_0^s}{1-\pi_0^s}\right) - \log\left(\frac{\pi_1^s}{1-\pi_1^s}\right)$$

Note that specifying $\underbrace{\lambda^s \text{ and } \pi_0^s}_{\text{need priors on these}}$ determines π_1^s.

Summary of 13 Selected Studies (again)

Study Summary Data

		Con			Treat		
Study	NRS	Days	$O_1 \equiv$	NRS	Days	$O_0 \equiv$	Odds Ratios
s	x_0^s	n_0^s	$\frac{x_0^s}{n_0^s - x_0^s}$	x_1^s	n_1^s	$\frac{x_1^s}{n_1^s - x_1^s}$	O_0/O_1
1	55	97	1.310	67	101	1.971	0.665
2	137	282	0.945	187	306	1.571	0.601
3	505	927	1.197	590	915	1.815	0.659
4	62	123	1.016	74	140	1.121	0.907
5	118	236	1.000	99	239	0.707	0.707
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The Bayesian Hierarchical Model (bottom-to-top)

Data models:

$$\text{Control: } x_0^s \sim \text{binomial}(\pi_0^s, n_0^s);$$
$$\text{Treatment: } x_1^s \sim \text{binomial}(\pi_1^s, n_1^s).$$

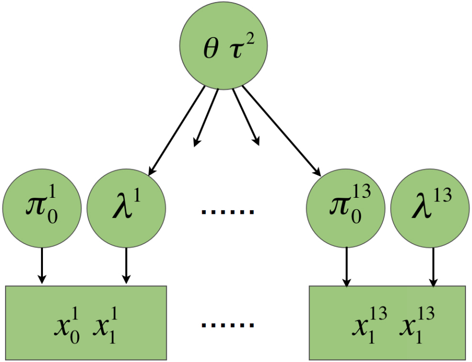
Study prior models:

$$\pi_0^s \sim \text{beta}(1, 1);$$
$$\lambda^s|\theta \sim N(\theta, \tau^2).$$

Population prior models:

$$\theta \sim N(0, 100);$$
$$\tau^2 \sim \text{inverse gamma}(3, 0.1).$$

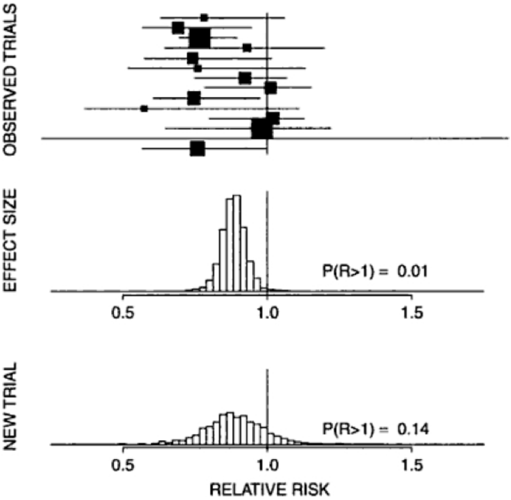
Graphical Hierarchical Model (top-to-bottom)



- Circles represent unknown parameters; rectangles represent data; arrows represent dependencies[4]
- The baseline NSR probabilities (π_0^s) are independent across studies

[4] Adapted from Parmigiani (2002), p. 131.

Posterior and Posterior Predictive Results



(See Parmigiani (2002), p. 137)

Posterior and Posterior Predictive Results, Cont.

- These graphs summarize inference for relative risk, $RR = \pi_0^s / \pi_1^s$; that is, the ratio of NRS rates in the control to the treatment group.
- Here, $RR < 1$ implies greater risk of NRS under BL treatment.
- Results in the previous graphs:
  - Top: Squares are proportional to study size and are centered at the empirical RR. Horizontal line length is a 98% posterior probability interval for RR.
  - Middle: This is the overall posterior for RR, assuming a baseline rate of 50%.
  - Bottom: Predictive distribution of RR, again assuming a baseline rate of 50%.

Prior Specification in Bayesian Meta Analysis

- Not easy!
- Empirical Bayesian methods (not "fully Bayesian").
- Large and growing literature on prior specification.
- Based on prior data:
  - Simple maximum likelihood or method of moments estimators
  - Power priors–when prior data yields overly precise priors
  - Conditional means priors–useful in generalized linear models
  - Use diffuse priors, combine with prior data to obtain posterior–use the latter as a posterior for the current meta analysis
- Subject matter expertise very important.
- Be mindful of induced priors and other subtleties...

Example: Meta-analysis on migraine headaches

- Two classes of treatments: beta-blockers & calcium-channel blockers (Parmigiani, 2002).
- Hierarchical model w/ multiple variance components:
  - variation of treatments within a class of treatments (τ_1^2),
  - variation between treatment classes (τ_2^2), and
  - variation of the overall treatment mean (τ_3^2).
- Hierarchical structure induces a correlation among treatment effects & *and orders them*:

$$\underbrace{\rho_0 = \frac{\tau_2^2 + \tau_3^2}{\tau_1^2 + \tau_2^2 + \tau_3^2}}_{\text{within same class}} \geq \underbrace{\rho_1 = \frac{\tau_3^2}{\tau_1^2 + \tau_2^2 + \tau_3^2}}_{\text{in different classes}}$$

- How to model this in the prior? Not obvious...

But not hard either...

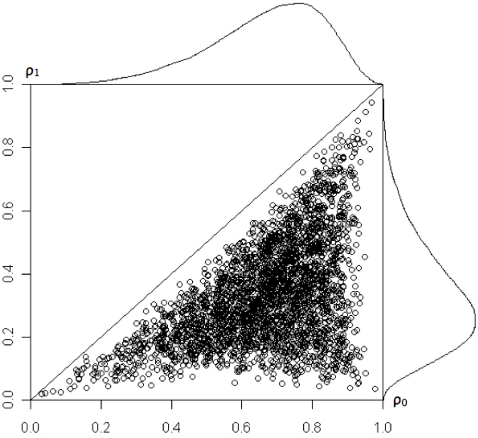


Figure: $\tau_j^2 \sim$ inverse-gamma$(3, 2/3)$ yields this joint prior. Note the induced marginal priors for the correlations. The details are technical; the point is that it can be done.

Bayesian Meta Analysis: Advantages

- Conclusions are probability statements; e.g., probability that the mean change from baseline exceeds zero.
- Conflict between fixed- and random-effects meta analysis overcome by modeling between-trial variability.
- The "random effects" distribution need not be normal–very flexible modeling.
- We can predict future results using the posterior predictive distribution.
  - Useful for sample size studies for confirmatory clinical trials.
  - Serves as a source of prior distributions for a future study.
- Modeling unobserved aspects of data generation and reporting processes. This includes publication bias, missing covariates, misclassification, measurement error, etc.

Bayesian Meta Analysis: Challenges & Disadvantages

- Choice of family for prior distributions and eliciting/fitting priors is typically very difficult.
- Prior-to-posterior sensitivity: different prior distributions can lead to very different conclusions. A sensitivity analysis is essential.
- There is no conventional measure of significance, such as with P-values. That is, "$\alpha = 0.05$" is not a common criteria. (But use of such criteria in frequentist methods is itself problematic–see papers by Goodman.)
- Communication of results: The "consumer" of Bayesian meta analysis conclusions will probably not be familiar with Bayesian methods. (But results are in terms of probabilities, so interpretation is easy.)

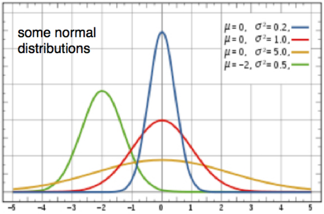
Conclusion: Bayes & Meta Analysis–A Natural Fit

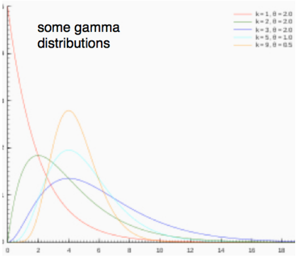
- "Meta-analysis is a natural for the Bayesian" (D.V. Lindley)
- "The Bayesian paradigm is synonymous with meta-analysis. For both, the goal is to incorporate all information to predict with as much accuracy as possible some future event, and the uncertainty associated with it, and to present this prediction in a manner that leads to coherent decisions."
  (D. Stangl & D. Berry)
- Updating through Bayes' theorem "...is common in all applications of Bayesian methods and demonstrates how all Bayesian analysis can be seen as meta-analysis."
  (D. Stangl & D. Berry)

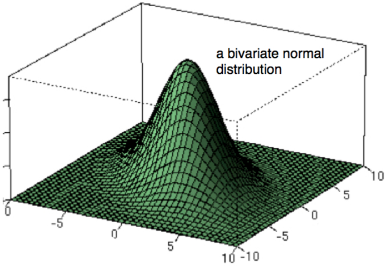
Appendix: Notation for Distributions

- $N(\mu, \sigma^2)$ means a normal distribution with mean μ and variance σ^2.
- Beta(α, β) means a beta distribution with shape parameters α and β. (Mean $\alpha/(\alpha+\beta)$; variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.)
- $U(a, b)$ means a uniform ("flat") distribution over the interval from the numbers a to b. Same as Beta(1,1). (Mean $(a+b)/2$; variance $(b-a)^2/4$.)
- gamma(κ, θ) is a gamma distribution with shape parameters κ and θ. (Mean $\kappa\theta$; variance $\kappa\theta^2$).
- inverse-gamma(u, v) means the reciprocal of a random variable with this distribution is gamma(u, v).

Appendix: Examples of Distribution Shapes (Wikipedia!)







Recommend Reading: Bayesian Statistics for Non-Statisticians

- Berger & Berry (1988) Statistical analysis and the illusion of objectivity, *American Scientist*, **76**, 159-165.
- Berry (2004) Bayesian statistics & the efficiency and ethics of clinical trials, *Statistical Science*, **19**, 175-187.
- Goodman (1999) Toward evidence-based medical statistics. 1: The p-value fallacy, *Ann Internal Med*, **130**, 995-1004.
- Goodman (1999) Toward evidence-based medical statistics. 2: The Bayes factor, *Ann Internal Med*, **130**, 1005-1013.
- Goodman (2001) Of p-values and Bayes: a modest proposal, *Epidemiology*, **12**, 295-297.
- O'Hagan and Luce (2003) *A Primer on Bayesian Statistics in Health Economics and Outcomes Research*, available as a pdf file at http://www.shef.ac.uk/content/1/c6/02/55/92/primer.pdf. (Highly recommended.)
- Spiegelalter (2004) Incorporating Bayesian ideas into health-care evaluation, *Statistical Science*, **19**, 156-174.

Recommend Reading: General Technical Sources

- Carlin & Louis (2009) *Bayes & Empirical Bayes Methods for Data Analysis*, 3rd Ed.
- Christensen et al. (2011) *Bayesian Ideas & Data Analysis*
- FDA (2010) Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials
- Gelman et al. (2004) *Bayesian Data Analysis*, 2nd Ed.
- Parmigiani (2002) *Modeling in Medical Decision Making* (Ch. 4 is on meta analysis–highly recommended!)
- Spiegelhalter, Abrams, & Myles (2004) *Bayesian Approaches to Clinical Trails & Health-Care Evaluation* (Ch. 8 is on meta analysis)
- Stangl & Berry (2000) *Meta-Analysis in Medicine & Health Policy*
- WinBugs website list: www.mrc-bsu.cam.ac.uk/bugs/references/bugs-uses-abstracts.shtml (This site refers to several papers on Bayesian meta analysis.)