DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

INTRODUCTION TO BAYESIAN METHODS IN BIOMEDICAL RESEARCH

Lurdes Y.T. Inoue, Ph.D.
Professor of Biostatistics

Applications of the Bayesian approach in biomedical research

American Journal of Epidemiology
Vol. 179, No. 3
DOI: 10.1093/aje/tet258
Advance Access publication: November 10, 2013

Practice of Epidemiology

Replication of Breast Cancer Susceptibility Loci in Whites and African Americans Using a Bayesian Approach

Katie M. O'Brien*, Stephen R. Cole, Charles Poole, Jeannette T. Bensen, Amy H. Herring, Lawrence S. Engel, and Robert C. Millikan†

SNPs. Our results demonstrate the utility of Bayesian methods in genetic epidemiology and provide support for their application in small, etiologically driven investigations.

Medical Decision Making
http://mdm.sagepub.com/

A Bayesian Approach to Aid in Formulary Decision Making: Incorporating Institution-Specific Cost-Effectiveness Data with Clinical Trial Results

Shelby D. Reed, Peter W. Dillingham, Andrew H. Briggs, David L. Veenstra and Sean D. Sullivan
Med Decis Making 2003 23: 252
DOI: 10.1177/0272989X03023003007

Then, we adopted a Bayesian, hierarchical, random-effects model to integrate site-specific and clinical trial data. We ap-

J Community Health (2011) 36:819–830
DOI 10.1007/s10900-011-9380-8

ORIGINAL PAPER

Comparing Child Health, Access to Care, and Utilization of Health Services Between Ohio Appalachia's River and Non-River Bordering Counties

Laureen H. Smith · Christopher Holloman

do not border the river. A secondary analysis of the 28 Appalachian counties from Ohio's 88 counties included in the 2008 Ohio Family Health Survey was conducted using a Bayesian Hierarchical Modeling strategy. Descriptive

Environ Resource Econ (2011) 49:597–624
DOI 10.1007/s10640-011-9456-z

The Effect of Risk Context on the Value of a Statistical Life: a Bayesian Meta-model

Thijs Dekker · Roy Brouwer · Marjan Hofkes · Klaus Moeltner

ACADEMIC EMERGENCY MEDICINE 2008; 15:466–475 © 2008 by the Society for Academic Emergency Medicine

Bayesian Logistic Injury Severity Score:
A Method for Predicting Mortality Using
International Classification of Disease-9 Codes

Randall S. Burd, MD, PhD, Ming Ouyang, PhD, David Madigan, PhD

Methods: The authors used Bayesian logistic regression to train and test models for predicting mortality based on injury ICD-9 codes (2,210 codes) and injury codes with two-way interactions (243,037 codes and interactions) using data from the National Trauma Data Bank (NTDB). They evaluated discrimination

American Journal of Epidemiology
Copyright © 1995 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

Vol. 141, No. 3
Printed in U.S.A.

Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard

Lawrence Joseph,[1–3] Theresa W. Gyorkos,[1,2,4] and Louis Coupal[2,3]

It is common in population screening surveys or in the investigation of new diagnostic tests to have results from one or more tests investigating the same condition or disease, none of which can be considered a gold standard. For example, two methods often used in population-based surveys for estimating the prevalence of a parasitic or other infection are stool examinations and serologic testing. However, it is known that results from stool examinations generally underestimate the prevalence, while serology generally results in overestimation. Using a Bayesian approach, simultaneous inferences about the population prevalence and the sensitivity, specificity, and positive and negative predictive values of each diagnostic test are possible. The methods presented here can be applied to each test separately or to two or more tests combined. Marginal posterior densities of all parameters are estimated using the Gibbs sampler. The techniques are applied to the estimation of the prevalence of *Strongyloides* infection and to the investigation of the diagnostic test properties of stool examinations and serologic testing, using data from a survey of all Cambodian refugees who arrived in Montreal, Canada, during an 8-month period. *Am J Epidemiol* 1995;141:263–72.

Bayes theorem; diagnostic tests, routine; epidemiologic methods; models, statistical; Monte Carlo method; prevalence; sensitivity and specificity

CASE STUDY

Causal Agent

The nematode (roundworm) *Strongyloides stercoralis*. Other Strongyloides include *S. fülleborni*, which infects chimpanzees and baboons and may produce limited infections in humans.

CASE STUDY

Life Cycle

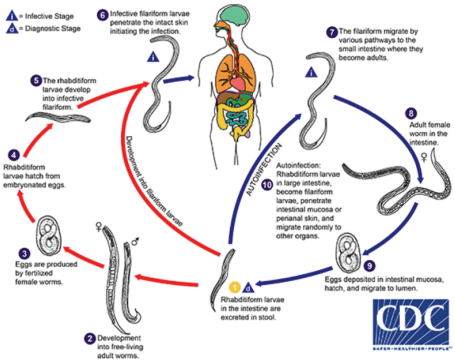


TABLE 1. Results of serologic and stool testing for *Strongyloides* infection on 162 Cambodian refugees arriving in Montreal, Canada, between July 1982 and February 1983

		Stool examination +	Stool examination −	
Serology	+	38	87	125
	−	2	35	37
		40	122	162

- **Goals:**
  - Estimate disease prevalence
  - Estimate sensitivity and specificity of each individual test
  - Estimate sensitivity and specificity of the combined tests
- **Challenge:**
  - No GOLD STANDARD evaluated in the study!

Additional Information

lack of a gold standard for the detection of most parasitic infections means that the properties of these tests are not known with high accuracy. In consultation with a panel of experts from the McGill Centre for Tropical Diseases, we determined equally tailed 95 percent probability intervals (i.e., 2.5 percent in each tail) for the sensitivity and specificity of each test (see table 5). These were derived from a review of the relevant literature and clinical opinion (21–28).

Bayesian Methods

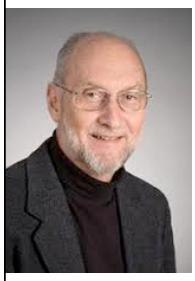
"the explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a health-care evaluation"

(Spiegelhalter, Abrams, Myles, 2004)

Statistical Science
1990, Vol. 5, No. 3, 269–298

Biostatistics and Bayes

Norman Breslow



Abstract. Attitudes of biostatisticians toward implementation of the Bayesian paradigm have changed during the past decade due to the increased availability of computational tools for realistic problems. Empirical Bayes' methods, already widely used in the analysis of longitudinal data, promise to improve cancer incidence maps by accounting for overdispersion and spatial correlation. Hierarchical Bayes' methods offer a natural framework in which to demonstrate the bioequivalence of pharmacologic compounds. Their use for quantitative risk assessment and carcinogenesis bioassay is more controversial, however, due to uncertainty regarding specification of informative priors. Bayesian methods simplify the analysis of data from sequential clinical trials and avoid certain paradoxes of frequentist inference. They offer a natural setting for the synthesis of expert opinion in deciding policy matters. Both frequentist and Bayes' methods have a place in biostatistical practice.

Message is out!

- Science
- The Economist
- The New York Times
  Adding Art to the
  Rigor of Statistical Science
  (Arts & Ideas Section)


Bayes Offers a 'New' Way to Make Sense of Numbers


In praise of Bayes


Arts & Ideas
Adding Art to the Rigor Of Statistical Science

Application Areas include
- Medicine
- Genetics
- Pharmacology
- Epidemiology
- Health services
- Environmental sciences

…

And increasing due to modeling flexibility, computational resources, etc…

Bayesian Software

(Disclaimer: Not intended to provide a complete list of available Bayesian software)

- BUGS/Winbugs/Openbugs/JAGS (complex models using MCMC methods)
- BOA/CODA (convergence diagnostics and output analysis)
- BRCAPRO (genetic counseling of women at risk for breast and ovarian cancer)

…

R-Packages:

- http://cran.r-project.org/web/views/Bayesian.html

  - Download Rstudio: https://www.rstudio.com/products/Rstudio/
  - Download and install R in your computer: http://cran.fhcrc.org/

  Within R session:
  - Install packages with
    - install.packages("mypackage")
  - Load library with
    - library(mypackage)

Primary packages we will use

- LearnBayes
- INLA
  - Download at: http://www.r-inla.org/download

  install.packages("INLA", repos=https://www.math.ntnu.no/inla/R/stable)
  library(INLA)
- arm
- rjags (alternative choices R2jags, runjags)

Basic Concepts/Review

- Probability & Interpretation
- Random Variables
- Likelihood Function
- Traditional Approach to Inference

Diagnostic Testing

- In the presence of a "gold standard"
  - Consider a new diagnostic test

		Disease	No Disease
Test	Positive	a (true positives)	b (false positives)
	Negative	c (false negatives)	d (true negatives)

- Events:
  - A: {test positive}
  - B: {disease} ⟶ P(B): disease prevalence

Diagnostic Testing

Test	Disease	No Disease
Positive	a	b
Negative	c	d

- **Sensitivity:** the ability of the test to identify correctly those who have the disease among all individuals with the disease

$$Sensitivity: P(A \mid B) = \frac{a}{a+c}$$

- **Specificity**: the ability of the test to identify correctly those who do not have the disease among those free from the disease

$$Specificity: P(A^c \mid B^c) = \frac{d}{b+d}$$

- These are test characteristics.

Diagnostic Testing

Test	Disease	No Disease
Positive	a	b
Negative	c	d

- **Positive predictive value (PPV):** The proportion of patients have the disease among those who tested positive

$$PPV: P(B \mid A) = \frac{a}{a+b}$$

- **Negative predictive value (NPV)**: The proportion of patients are actually free of the disease among those who tested negative

$$NPV: P(B^c \mid A^c) = \frac{d}{c+d}$$

Interpretations of Probability

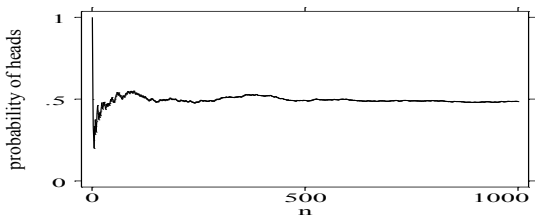
- Classical: If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a characteristic of interest, A, the probability of the occurrence of E is P(E) = m/N.

  Example: Flip a coin.

  What is the probability of getting a head?

- Frequentist: If some experiment is repeated a large number of times n and if some resulting event with the characteristic E occurs m times, the relative frequency of occurrence of E is approximately equal to the probability of E, that is, P(E) = m/n

Example: Around 1900, Karl Pearson tossed a coin 24,000 times and recorded 12,012 heads, giving a proportion of 0.5005.



Interpretations of Probability: Subjective

- Your degree of uncertainty.

Example: Will you pass a class?

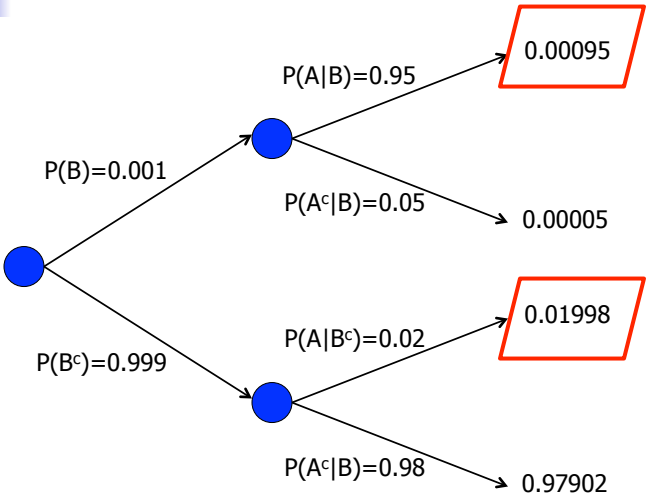
You will take the class (hopefully!) only once; even if you retake the class next year, you won't be taking it under the same conditions! You'll have a different instructor, a different set of courses, and possibly different working conditions!

Diagnostic Testing

- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/1000, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/1000
  Sensitivity = P(A|B) = 0.95
  Specificity = $P(A^c|B^c) = 0.98 = 1-P(A|B^c)$ = 1-False Positive

Diagnostic Testing



Conditional on positive test result:

PPV= 0.00095/(0.0095+0.01998)=0.045

Diagnostic Testing

- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/1000, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/1000
  Sensitivity = P(A|B) = 0.95
  Specificity = $P(A^c|B^c)$ = 0.98 = 1-$P(A|B^c)$ = 1-False Positive

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

Bayes Rule!

$$= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999} = \frac{0.00095}{0.02093} = 0.045$$

Positive Predictive Value

Diagnostic Testing

- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/100, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/100
  Sensitivity = P(A|B) = 0.95
  Specificity = $P(A^c|B^c)$ = 0.98 = 1-$P(A|B^c)$

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

Bayes Rule!

$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.02 \times 0.99} = 0.324$$

Positive Predictive Value

Diagnostic Testing

- Question: *How should the test result change our belief about the probability of disease?*
  - Our intuition is poor when processing probabilistic evidence, i.e., when updating our probability in the presence of new evidence. Bayes rule shows exactly how to do this!
  - **The disease prevalence** (0.001) can be thought of as our **prior** probability that the individual has the disease.
  - Observing a positive result (i.e. data) changes this probability to 0.045 for the tested individual. This is our updated or ***posterior*** *probability* that the individual has the disease.
  - The posterior probability depends on the test's operating characteristics (e.g. sensitivity/specificity, test results and prevalence).

Diagnostic Testing

- Questions:
  - *Having observed a positive test result for a subject, what is the probability that the next subject also has a positive test result?*
  - *How would the new test result change the current belief about the probability of disease?*

Guiding principle*: Today's posterior is tomorrow's prior!*

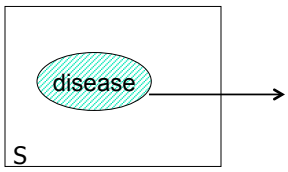
$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

$$= \frac{0.95 \times 0.045}{0.95 \times 0.045 + 0.02 \times (1 - 0.045)} = \frac{0.04275}{0.06185} = 0.691$$

What is a probability model?

- Random variable:
  - "Rule" that assigns a "value" to each point of the sample space

Example



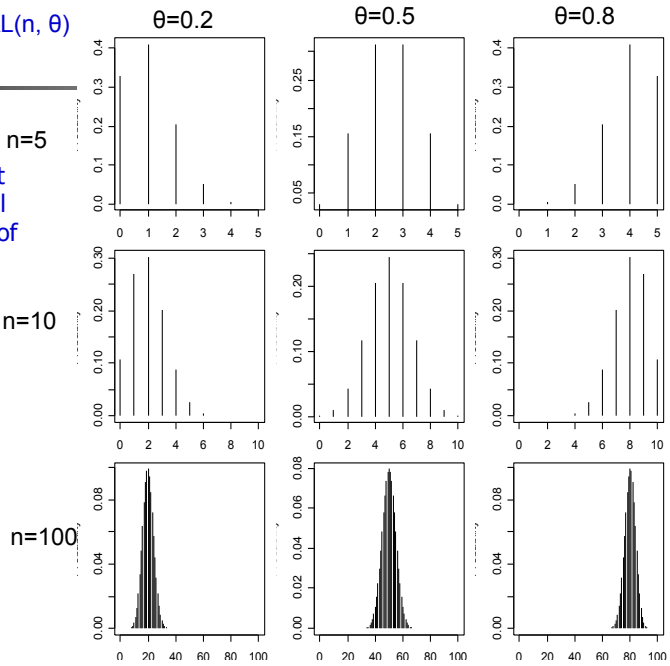
$$X = \begin{cases} 1, & \text{if subject has disease} \\ 0, & \text{otherwise} \end{cases}$$

- Probability model (of a random variable):
  - Defines what values the variable can take and how to assign probabilities to those values.

    Example: X ~ Bernoulli(p); p is the probability of disease

BINOMIAL(n, θ)

Number of times that a particular event will occur in a sequence of n independent observations with the same probability of occurrence

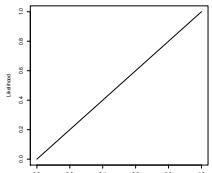


What is a likelihood function?

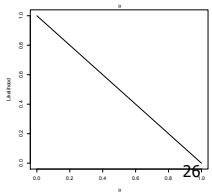
- A **likelihood function** (or simply the likelihood) is a function of the parameters of a probability model given the outcomes.
  - The *likelihood* of θ, given outcome y, is equal to the *probability* of that observed outcome given θ.

What is a likelihood function?

- Bernoulli model:
  - Random variable Y takes on two possible values: 0 or 1
    - $P(Y=1|\theta) = \theta$,
    - $P(Y=0|\theta) = 1-\theta$, where θ is a number in [0,1]
  - Likelihood function based on a Bernoulli observation:
    - Given that y=1, the likelihood function of θ is:
      - $L(\theta|y=1) = P(Y=1|\theta)=\theta$

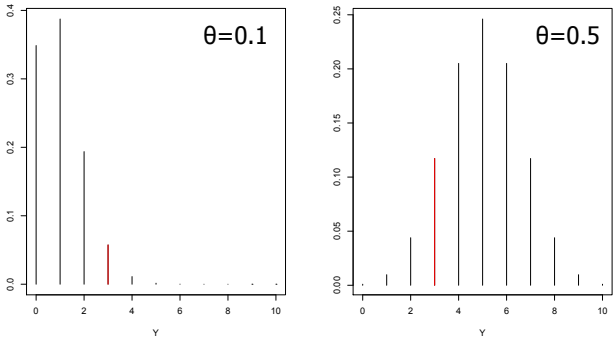


    - Given that y=0, the likelihood function of θ is:
      - $L(\theta|y=0) = P(Y=1|\theta)=1-\theta$



What is a likelihood function?

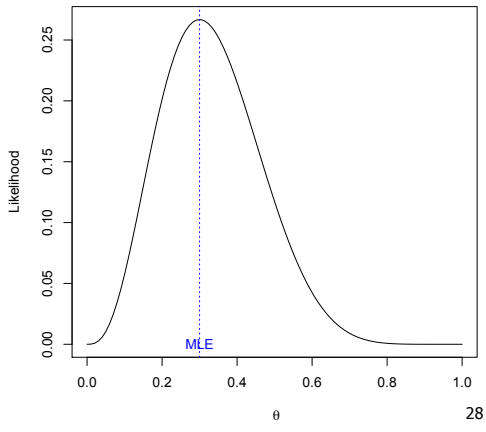
- Binomial Model
  - Test results in a random sample of 10 disease subjects: (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
  - Probability model for number of positive tests:
    - Y ~ Binomial(10, θ)



What is a likelihood function?

- Likelihood function: $L(\theta \mid Y) = \binom{10}{3} \theta^3 (1-\theta)^7$

What is the value of θ that maximizes the likelihood?



Traditional Approach to Inference

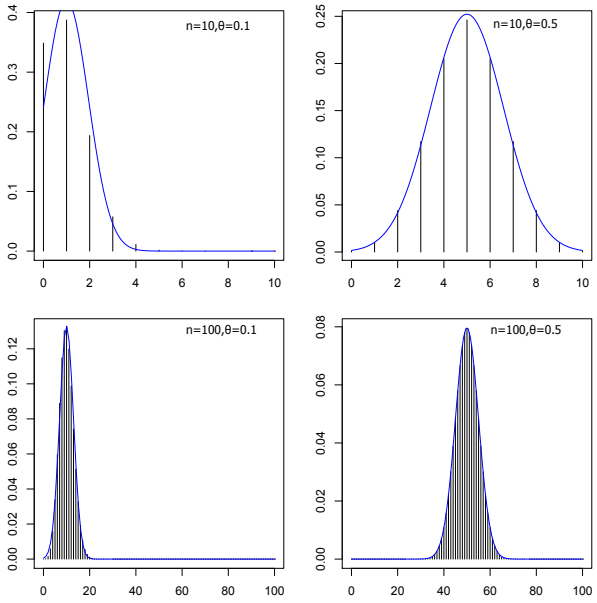
Under certain regularity conditions and for large samples:

$$\hat{\theta}_{MLE} \dot{\sim} N\left(\theta, I^{-1}(\theta)\right), \quad \text{where } I(\theta) = E_{Y|\theta}\left[-\frac{\partial^2 \log L(\theta \mid y)}{\partial \theta^2}\right]$$

Traditional Approach to Inference

Sampling Distribution

(Binomial and Normal Approximation)



Traditional Approach to Inference

H_0: θ = 0.5

H_1: θ ≠ 0.5

```
> binom.test(3,10,p=0.5)
        Exact binomial test
data:  3 and 10
number of successes = 3, number of trials = 10, p-value = 0.3438
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.06673951 0.65245285
sample estimates:
probability of success
                   0.3


> prop.test(3,10,p=0.5)
        1-sample proportions test with continuity correction
data:  3 out of 10, null probability 0.5
X-squared = 0.9, df = 1, p-value = 0.3428
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.08094782 0.64632928
sample estimates:
  p
0.3
```


Traditional Approach to Inference

- P-value interpretation?
  - Under the null hypothesis, the probability of observing an equal or more extreme number of test results is 34%.
  - *It is not the probability of the null hypothesis!*
- Confidence interval interpretation?
  - The confidence interval gives values of the population parameter for which the observed sample proportion is not statistically significant at the 5% level
  - *It does not give us the probability that the true parameter lies between the boundaries of the interval!*

Bayesian Approach to Inference

- Overview
- Prior Elicitation
- Prior Distributions
- Introduction to Bayesian Computation

Overview of the Bayesian approach

- Began with the work by Thomas Bayes who, in 1763, formalized what is now called Bayes Theorem.

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$$

where: $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$

Example: Diagnostic testing

- Data ➔ Result of test
- Parameter ➔ True disease status

Model

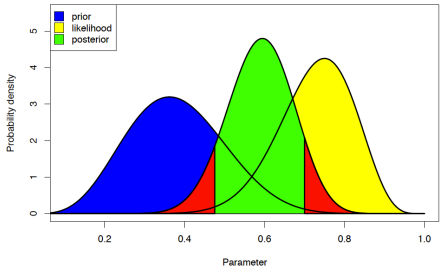
- Prevalence ➔ PRIOR PROB. OF DISEASE
- Sensitivity ➔ LIKELIHOOD of disease given positive test
- Specificity ➔ LIKELIHOOD of no disease given negative test

Bayes Theorem $P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$

- Positive Predictive Value ➔ POSTERIOR PROB. OF DISEASE GIVEN POSITIVE TEST
- Negative Predictive Value ➔ POSTERIOR PROB. OF NO DISEASE GIVEN NEGATIVE TEST

Overview of the Bayesian approach

- Moving towards a generic formulation:
  - Goal: learning about an unknown parameter θ (possibly a vector)
    - θ = true disease status
    - θ = hazard ratio
    - θ = probability that experimental treatment is better
    - θ = vector of regression coefficients
    - θ = missing data
    - etc...
  - Data: y (e.g. test result)
  - Input of analysis:
    - Prior distribution: $P(\theta)$
    - Probability Model: $P(y|\theta)$
    - Likelihood Function: $L(\theta|y) \propto P(y|\theta)$
  - Output of analysis:
    - Posterior distribution:



$$P(\theta \mid Y) = \frac{P(\theta)L(\theta \mid Y)}{\int P(\theta)L(\theta \mid Y)d\theta}$$

Overview of the Bayesian approach

- Inferences based on summaries of the posterior distribution
  - Point estimates:
    - Mean/Median/Mode
  - Interval estimates:
    - One-sided credible intervals
    - Two-sided credible intervals
      - Equi-tail area
      - Narrowest interval
        [HPD: highest posterior density intervals]

Choices of summary measures justified with loss functions [decision theory].

Prior Distributions

- Quantifiable (prior) beliefs exist in medicine
  - "… it is generally unrealistic to hope for large treatment effects…"
  - "… it might be reasonable to hope that a new treatment for acute stroke or acute myocardial infarction could reduce recurrent stroke or death rates in hospital from 10% to 9% or 8%, but not to hope that it could halve in-hospital mortality"

(Peto and Baigent, 1998, BMJ)

Prior Distributions

- Key role in Bayesian analysis
- Choice of priors is based on judgments and a degree of subjectivity cannot be avoided
  - Prior is not unique!
  - Sensitivity analysis is crucial in assessing the impact of particular distributions on the conclusions.
- Can we turn informal prior knowledge into a mathematical prior distribution? How?

Childhood Polyarteritis nodosa

PLoS One. 2015 Mar 30;10(3):e0120981. doi: 10.1371/journal.pone.0120981. eCollection 2015.

Elicitation of expert prior opinion: application to the MYPAN trial in childhood polyarteritis nodosa.

Hampson LV[1], Whitehead J[1], Eleftheriou D[2], Tudur-Smith C[3], Jones R[4], Jayne D[5], Hickey H[6], Beresford MW[7], Bracaglia C[8], Caldas A[9], Cimaz R[10], Dehoorne J[11], Dolezalova P[12], Friswell M[13], Jelusic M[14], Marks SD[15], Martin N[16], McMahon AM[17], Peitz J[18], van Royen-Kerkhof A[19], Soylemezoglu O[20], Brogan PA[2].

Author information

Abstract

OBJECTIVES: Definitive sample sizes for clinical trials in rare diseases are usually infeasible. Bayesian methodology can be used to maximise what is learnt from clinical trials in these circumstances. We elicited expert prior opinion for a future Bayesian randomised controlled trial for a rare inflammatory paediatric disease, polyarteritis nodosa (MYPAN, Mycophenolate mofetil for polyarteritis nodosa).

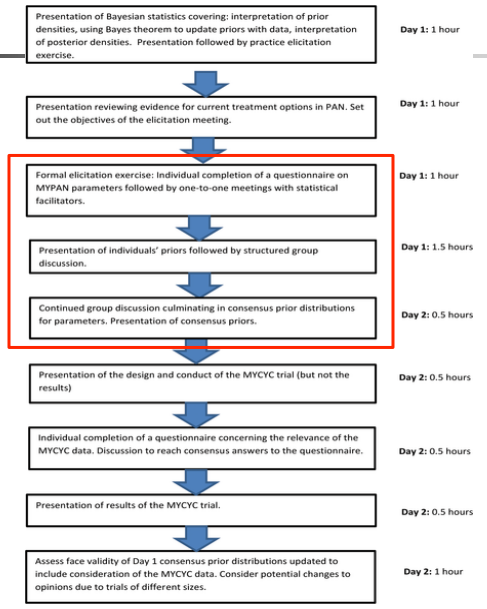
METHODS: A Bayesian prior elicitation meeting was convened. Opinion was sought on the probability that a patient in the MYPAN trial treated with cyclophosphamide would achieve disease remission within 6-months, and on the relative efficacies of mycophenolate mofetil and cyclophosphamide. Expert opinion was combined with previously unseen data from a recently completed randomised controlled trial in ANCA associated vasculitis.

RESULTS: A pan-European group of fifteen experts participated in the elicitation meeting. Consensus expert prior opinion was that the most likely rates of disease remission within 6 months on cyclophosphamide or mycophenolate mofetil were 74% and 71%, respectively. This prior opinion will now be taken forward and will be modified to formulate a Bayesian posterior opinion once the MYPAN trial data from 40 patients randomised 1:1 to either CYC or MMF become available.

CONCLUSIONS: We suggest that the methodological template we propose could be applied to trial design for other rare diseases.

CASE STUDY

ow diagram illustrating the sequence of activities undertaken during the MYPAN prior elicitation meeting and the time allocated to each activity.



Hampson LV, Whitehead J, Eleftheriou D, Tudur-Smith C, Jones R, et al. (2015) Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. PLoS ONE 10(3): e0120981. doi:10.1371/journal.pone.0120981
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0120981

PLOS ONE

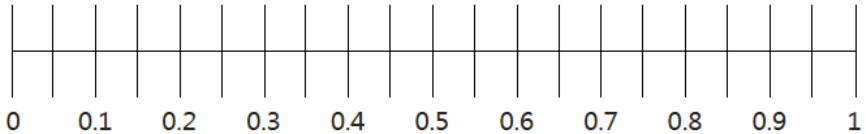
CASE STUDY

S1 File: Structured questionnaire designed to systematically ascertain prior opinion regarding outcomes for treatment with CYC and MMF

NAME:

Before any data are observed, please answer the following questions to specify your prior distributions.

Mark on the scales below your answers to the following questions (to the nearest 0.05).



Questionnaire

- Q1: What do you think the 6-month remission rate for children with PAN treated with cyclophosphamide (CYC) in combination with corticosteroids (steroids) is?
- Q2: Provide a proportion such that you are 75% sure that the true 6-month remission rate on CYC/steroids exceeds this value.

Because of the unpleasant side-effects of CYC, mycophenolate mofetil (MMF) might be considered the preferable treatment even if it is associated with a somewhat lower 6-month remission rate:

- Q3: What is the chance that the 6-month remission rate on MMF/steroids is higher than that on CYC/steroids?
- Q4: What is the chance that the 6-month remission rate on CYC/steroids exceeds that on MMF/steroids by more than 10%?

Please answer the following questions which will allow us to check the adequacy of your fitted prior distributions.

- Q5: What do you think the 6-month remission rate on MMF/steroids is?
- Q6: Provide a proportion such that you are 75% sure that the true 6-month remission rate on MMF/steroids exceeds this value.

S1 Table: Individual experts' final answers to Q1-Q4 and consensus answers agreed by the group before results from the MYCYC trial were revealed

Expert	Q1	Q2	Q3	Q4
1	0.65	0.45	0.63	0.05
2	0.85	0.60	0.35	0.20
3	0.80	0.55	0.10	0.50
4	0.85	0.65	0.20	0.40
5	0.70	0.60	0.20	0.20
6	0.80	0.80	0.15	0.10
7	0.75	0.50	0.10	0.15
8	0.75	0.55	0.30	0.20
9	0.70	0.60	0.20	0.10
10	0.70	0.60	0.25	0.25
11	0.75	0.55	0.30	0.20
12	0.70	0.50	0.10	0.30
13	0.75	0.40	0.20	0.15
14	0.80	0.55	0.20	0.35
15	0.80	0.60	0.20	0.30
Mean	**0.76**	**0.57**	**0.23**	**0.23**
Median	**0.75**	**0.55**	**0.20**	**0.20**
Consensus values[†]	**0.70**	**0.50**	**0.30**	**0.30**

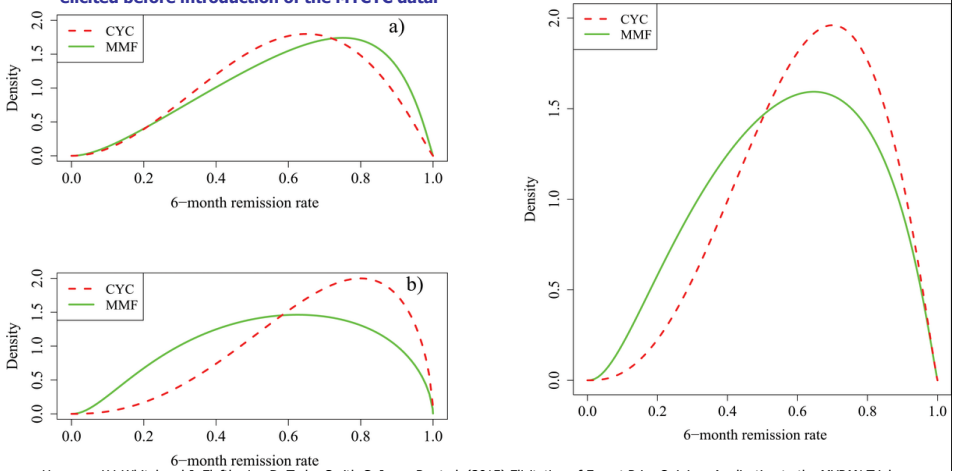
Q1: What do you think the 6-month remission rate for children with PAN treated with cyclophosphamide (CYC) in combination with corticosteroids (steroids) is?
Q2: Provide a proportion such that you are 75% sure that the true 6-month remission rate on CYC/steroids exceeds this value.
Q3: What is the chance that the 6-month remission rate on MMF/steroids is higher than that on CYC/steroids?
Q4: What is the chance that the 6-month remission rate on CYC/steroids exceeds that on MMF/steroids by more than 10%?

Consensus Prior

- Consensus to questions determined by vote.
  - Experts voted for the pair of answers to (Q1, Q2) which they thought best reflected their prior opinion for p_C.
  - Votes cast between pairs of answers (0.7, 0.5) and (0.75, 0.55), received 10 (67%) and 4 (27%) votes, respectively; one expert abstained.
  - Consensus answers were those voted for by the majority as reflecting their opinion.
- Consensus to (Q3, Q4) determined similarly
  - Experts votes between the following pairs of answers: (0.3, 0.3) and (0.3, 0.35) received 12 (80%) and 3 (20%) votes, respectively.

Fig 3. Expert prior opinion before introduction of the MYCYC data regarding 6-month remission rates using treatment with CYC or MMF for children with PAN.

Fig 1. Range of prior opinions elicited before introduction of the MYCYC data.



Hampson LV, Whitehead J, Eleftheriou D, Tudur-Smith C, Jones R, et al. (2015) Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. PLoS ONE 10(3): e0120981. doi:10.1371/journal.pone.0120981
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0120981

PLOS ONE

Prior elicitation

- Elicitation of prior distributions can be made from a number of people (for example, clinicians and patients)
  - Combined group (hierarchical) prior distribution
  - Consensus
  - Multiple prior distributions
    - Clinical prior: averages prior distributions elicited from experts
    - Vague prior: leads to a posterior distribution proportional to the likelihood
    - Skeptical prior: represents no treatment effect
    - Enthusiastic prior: represents large treatment effect

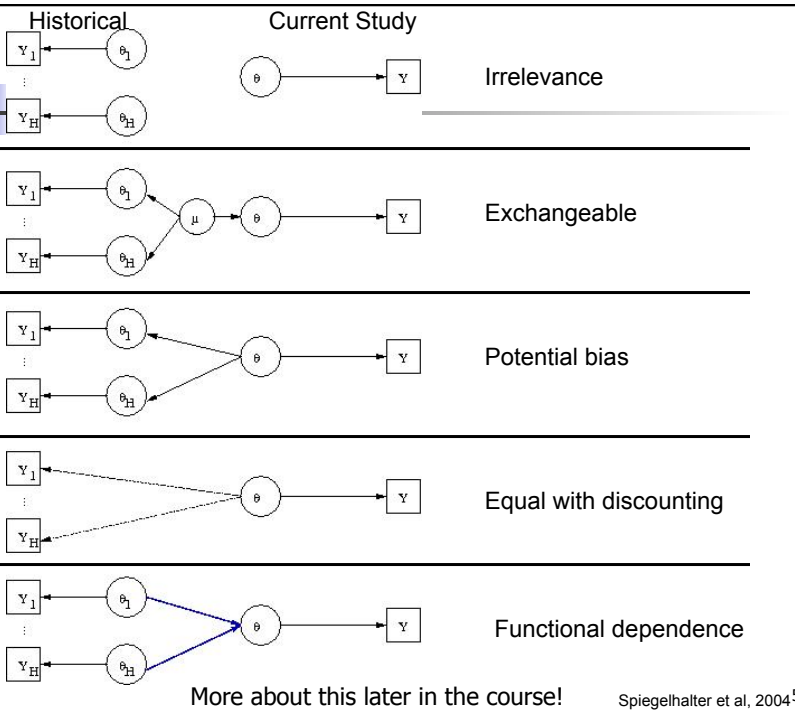
Prior elicitation

- General recommendations:
  - Interactive feedback: helps formulate probabilistic ideas and to reconcile inconsistencies
  - Scripted interview: uniformity in the elicitation process across experts
  - Review: the expert should have access to literature review
  - Percentile: Useful to consider 2.5^{th} and 97.5^{th} percentiles (95% probability intervals)

Prior elicitation

- Problem: how to turn informal opinions into a mathem prior distribution?
  - Summarizing historical evidence
    - Previous similar studies/trials can be used as the basis of a prior distribution
    - Several modeling approaches
      - Degrees of "similarity" between studies/trials
      - Possibility of bias

Note: These approaches are also used when considering historical controls in randomized trials, modeling for potential biases in observational studies and in pooling data for evidence synthesis (meta-analysis)



More about this later in the course!

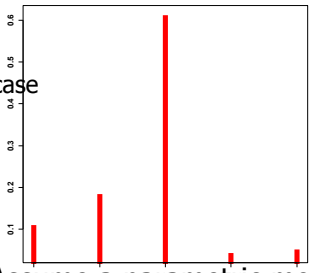
Spiegelhalter et al, 2004

Prior elicitation

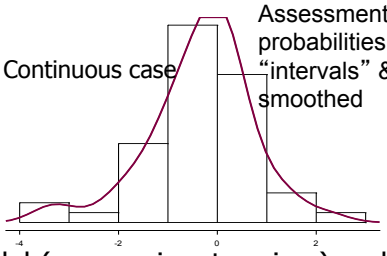
- Problem: how to turn informal opinions into a mathematical prior distribution?
  - Elicitation of subjective opinion

    - 'Histogram' approach

Discrete case



Continuous case



Assessment of probabilities in "intervals" & smoothed

  - Assume a parametric model (e.g. conjugate priors) and elicit quantities of interest

Prior Distributions

- Conjugate priors
- Non-informative
- Hierarchical priors
- Mixture priors

Prior Distributions

- Conjugate priors:
  - Let F denote a class of sampling distributions p(y|θ) and P a class of prior distributions for θ. Then P is conjugate for F

    $p(\theta|y) \in P$ for all $p(.|\theta) \in F$ and $p(.) \in P$

    [prior and the posterior distribution are of the same family].
  - Interpreted as "prior data"
  - Computational convenience

Likelihood	Prior	Posterior
$X\|\theta \sim \mathcal{N}(\theta, \sigma^2)$	$\theta \sim \mathcal{N}(\mu, \tau^2)$	$\theta\|X \sim \mathcal{N}(\frac{\tau^2}{\sigma^2+\tau^2}X + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$
$X\|\theta \sim \mathcal{B}(n, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X \sim \mathcal{B}e(\alpha + x, n - x + \beta)$
$X_1, \ldots, X_n\|\theta \sim \mathcal{P}(\theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{G}a(\sum_i X_i + \alpha, n + \beta).$
$X_1, \ldots, X_n\|\theta \sim \mathcal{NB}(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{B}e(\alpha + mn, \beta + \sum_{i=1}^n x_i)$
$X \sim \mathcal{G}(n/2, 2\theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(n/2 + \alpha, (x/2 + \beta^{-1})^{-1})$
$X_1, \ldots, X_n\|\theta \sim \mathcal{U}(0, \theta)$	$\theta \sim \mathcal{P}a(\theta_0, \alpha)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{P}a(\max\{\theta_0, x_1, \ldots, x_n\}\alpha + n)$
$X\|\theta \sim \mathcal{N}(\mu, \theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(\alpha + 1/2, \beta + (\mu - X)^2/2)$
$X\|\theta \sim \mathcal{G}a(\nu, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X \sim \mathcal{G}a(\alpha + \nu, \beta + x)$

Prior Distributions

- Non-informative:

  (reference prior, vague prior or flat prior)
  - Intended to provide "objective" analysis
    - Connections to Frequentist Inference!
  - Prior is "flat" relative to the likelihood function
    - Minimal impact on the posterior distribution of θ.
  - May be improper (does not "sum up" to 1)
    - DANGER: may lead to improper posteriors!!

Prior Distributions

- Discrete parameter:
  - Discrete uniform prior
    - Example:
      - Parameter = true hypothesis (null or alternative)
      - Prior: $P(H_0)=P(H_1)=0.5$
- Continuous parameter:
  - Jeffreys' prior

$$P(\theta) = |I(\theta)|^{1/2}, \text{where } I(\theta) = E\left[-\frac{\partial^2 \log P(Y \mid \theta)}{\partial \theta_i \partial \theta_j}\right] \text{(Fisher information)}$$

    - Idea: Fisher information measures the curvature of the log-likelihood. High curvature occurs whenever small changes in the parameter values are associated with large changes in the likelihood. Jeffreys' prior gives more weight to those parameter values, ensuring that the influence of the data and the prior essentially coincide
    - Invariant to transformations of θ

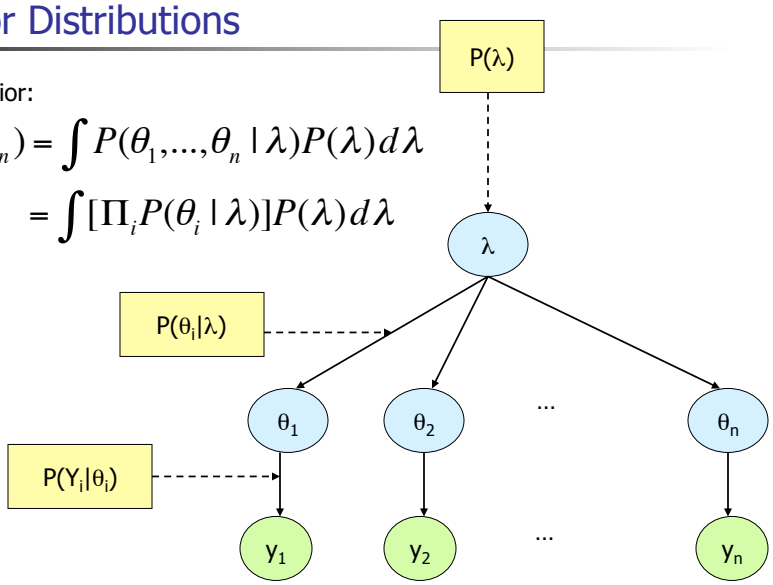
Prior Distributions

- Hierarchical priors:
  - Prior specification in phases
    - Structural division into stages
    - Quantitative (subjective) specification at each stage
- Borrowing strength:
  - improves precision for each parameter
- Nothing prevent us from going further into the hierarchy and adding stages.
  - Harder to interpret parameters in higher levels of the hierarchy
  - Common practice: non-informative priors at the higher levels (of course, "caveats" to such choices)

Prior Distributions

Hierarchical Prior:

$$P(\theta_1,\ldots,\theta_n) = \int P(\theta_1,\ldots,\theta_n \mid \lambda)P(\lambda)\,d\lambda$$
$$= \int [\Pi_i P(\theta_i \mid \lambda)]P(\lambda)\,d\lambda$$



Application: Meta-Analysis (more later!)

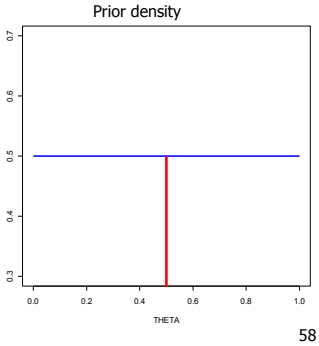
Prior Distributions

- Mixture Prior:
  - Example:

    Test results among 10 disease subjects:
    - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1) : 'successes'=3, 'failures'=7

H_0: θ = 0.5 versus H_1: θ ≠ 0.5

Priors for hypotheses:
- $P(H_0)=P(H_1)=0.5$
- Under alternative: θ ~ Beta(1,1)



- Prior can be re-written as a mixture:

$$P(\theta) = 0.5 \times I_{\{\theta=0.5\}} + 0.5 \times U(0,1)$$

American Journal of Epidemiology
Vol. 141, No. 3
Printed in U.S.A.

CASE STUDY

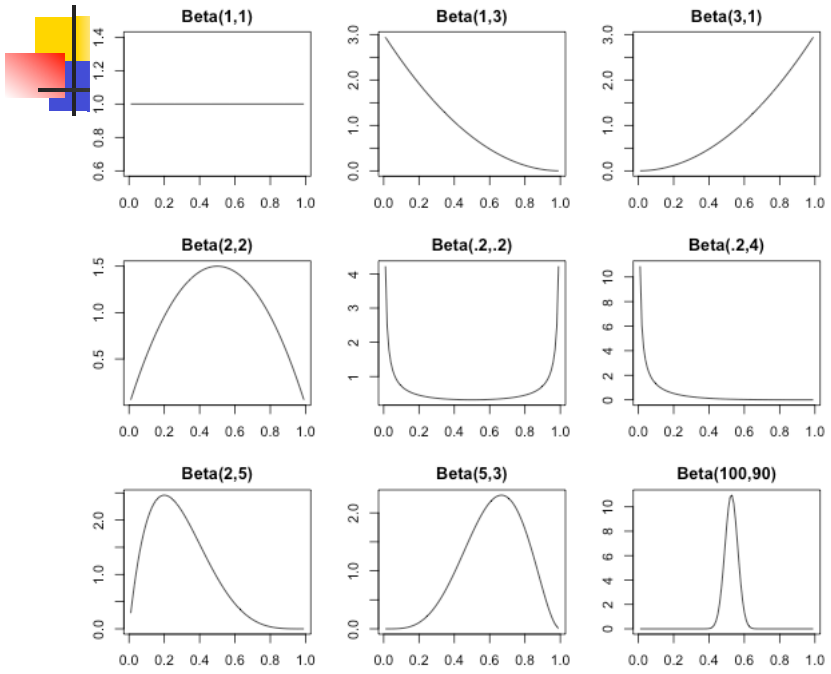
Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard

lack of a gold standard for the detection of most parasitic infections means that the properties of these tests are not known with high accuracy. In consultation with a panel of experts from the McGill Centre for Tropical Diseases, we determined equally tailed 95 percent probability intervals (i.e., 2.5 percent in each tail) for the sensitivity and specificity of each test (see table 5). These were derived from a review of the relevant literature and clinical opinion (21–28).

TABLE 5. Equally tailed 95% probability ranges and coefficients of the beta prior densities for the test parameters in the diagnosis of *Strongyloides* infection*

	Stool examination			Serology		
	Range (%)	Beta coefficients		Range (%)	Beta coefficients	
		α	β		α	β
Sensitivity	5–45	4.44	13.31	65–95	21.96	5.49
Specificity	90–100	71.25	3.75	35–100	4.1	1.76

* A uniform density over the range [0,1] ($\alpha=1$, $\beta=1$) was used for the prior distribution for the prevalence of *Strongyloides* in the refugee population.



Translating the information into a prior distribution

The particular beta prior density for each test parameter was selected by matching the center of the range with the mean of the beta distribution, given by $\alpha/(\alpha+\beta)$, and matching the standard deviation of the beta distribution, given by

$$\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}},$$

with one quarter of the total range. These two conditions uniquely define α and β. An alternative approach is to match the end points of the given ranges to beta distributions with similar 95 percent probability intervals. The coefficients obtained from these two approaches usually give very similar prior distributions. One way to consider a beta(α,β) distribution is to equate it with the information contained in a prior sample of $(\alpha + \beta)$ subjects, α of whom were positive. The sum $(\alpha + \beta)$ is often referred to as the "sample size equivalent" of the prior information (18).

	Range (%)	Beta coefficients	
		α	β
Sensitivity	5–45	4.44	13.31

Beta distribution obtained by solving these equations:

$$\frac{\alpha}{\alpha+\beta} = \frac{(.45+.05)}{2} = .25$$

$$\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \frac{1}{4}(.45-.05) = .10$$

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

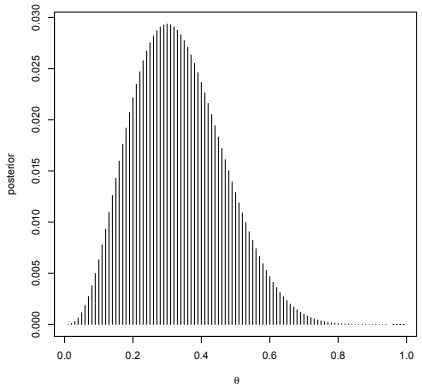
- Bayesian inference can be achieved by approximating the continuous θ with a (dense) grid of discrete values.
- A disadvantage of this approach is that the approximation is only as good as the grid is.
- An advantage of this approach is that it provides flexibility in the choice of prior distributions.
- We will illustrate this approach using
  - "brute-force" method (simple application of Bayes rule) or,
  - R package (LearnBayes)

Introduction to Bayesian Computation

- Test results of 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)
- Parameter of interest:
  - Probability of disease

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

```
# Method 1: Brute-force ---------------------------------------
#-- Prior 1:
prior <- rep(1/99, 99)

#-- likelihood times prior
product <- dbinom(x=data[1], size=sum(data), prob=theta)*prior

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot posterior distribution
plot(theta, posterior, type='h', xlab=expression(~theta))

#-- posterior mean
mean.post <- sum(theta*posterior)

#-- cumulative posterior distribution
cumulative.post <- cumsum(posterior)

#-- median (approximate)
median.post <- theta[max(which(cumulative.post <=0.50))]
```

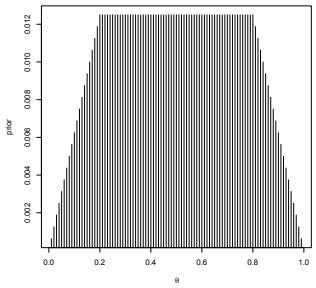


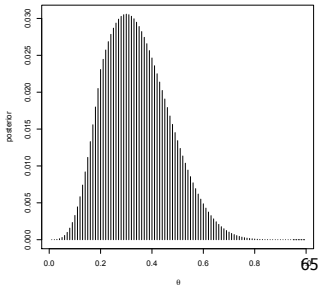
$$P(\theta_i \mid Y) = \frac{P(\theta_i)P(Y \mid \theta_i)}{\sum_j P(\theta_j)P(Y \mid \theta_j)}$$

```
> mean.post
[1] 0.3333333

> median.post
[1] 0.31
```

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

```
# Method 1: Brute-force ----------------------------------------

#-- Prior 2:
f <- theta*(theta<=0.2) +(1-theta)*(theta>=0.8) + 0.2*(theta>0.2 & theta < 0.8)
prior <- f/sum(f)

#-- likelihood times prior
product <- dbinom(x=data[1], size=sum(data), prob=theta)*prior

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot of prior distribution
plot(theta, prior, type='h', xlab=expression(~theta))

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot posterior distribution
plot(theta, posterior, type='h', xlab=expression(~theta))

#-- posterior mean
mean.post <- sum(theta*posterior)

#-- cumulative posterior distribution
cumulative.post <- cumsum(posterior)

#-- median (approximate)
median.post <- theta[max(which(cumulative.post <=0.50))]
```

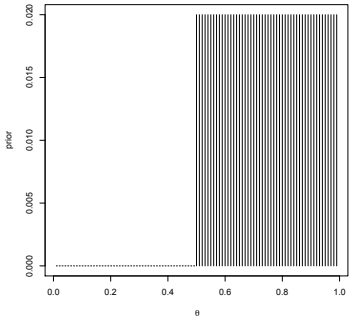


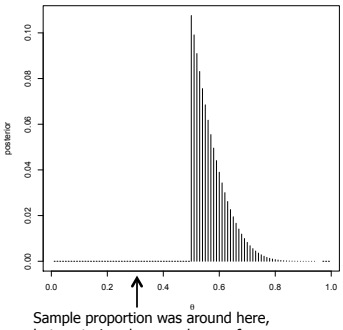



```
> mean.post
[1] 0.342133

> median.post
[1] 0.32
```


Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

Suppose a prior which places probability zero for θ < 0.5 and uniform otherwise





Sample proportion was around here, but posterior places prob. zero for values < .5!

Prior distributions

- Be careful!
  - **Cromwell's rule**:
    - "If a coherent Bayesian attaches a prior probability of zero to the hypothesis that the Moon is made of green cheese, then even whole armies of astronauts coming back bearing green cheese cannot convince him otherwise" (Lindley, 1985)
    - In other words, by placing a prior probability of zero, then there is no learning with data!

Overview of the Bayesian approach

- Likelihood function: $L(\theta \mid Y) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$

  where y: number of successes
  n: sample size

- Prior?
  - Let's consider a prior with a functional form that resembles that of the likelihood function
    - Prior should be of the form $\theta^{a'}(1-\theta)^{b'}$
    - It turns out that such a prior for θ is a Beta

Cool fact: multiply likelihood and the prior and you'll again get a function of the same form as the prior…

Overview of the Bayesian approach

- Likelihood function: $L(\theta \mid Y) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$
- Prior: $\theta \sim Beta(a,b)$ and $P(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$
  - a: "prior" successes
  - b: "prior" failures
- Posterior (via Bayes Theorem):

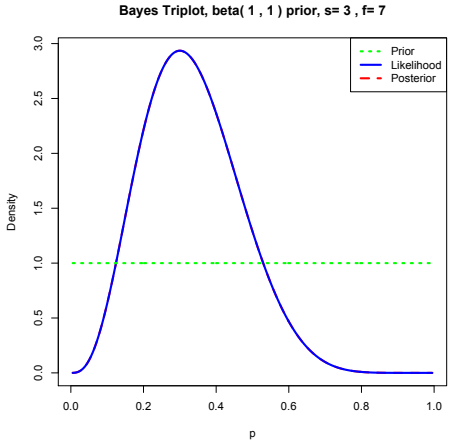
$$P(\theta \mid Y) \propto \theta^{y}(1-\theta)^{n-y}\theta^{a-1}(1-\theta)^{b-1}$$
$$\propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}$$

$$(\theta \mid Y) \sim Beta(a+y, b+n-y)$$

Introduction to Bayesian Computation: conjugate models

Bayesian Inference for a Proportion Using R:

```
library(LearnBayes)
triplot(prior=c(1,1),data=c(3,7))
```

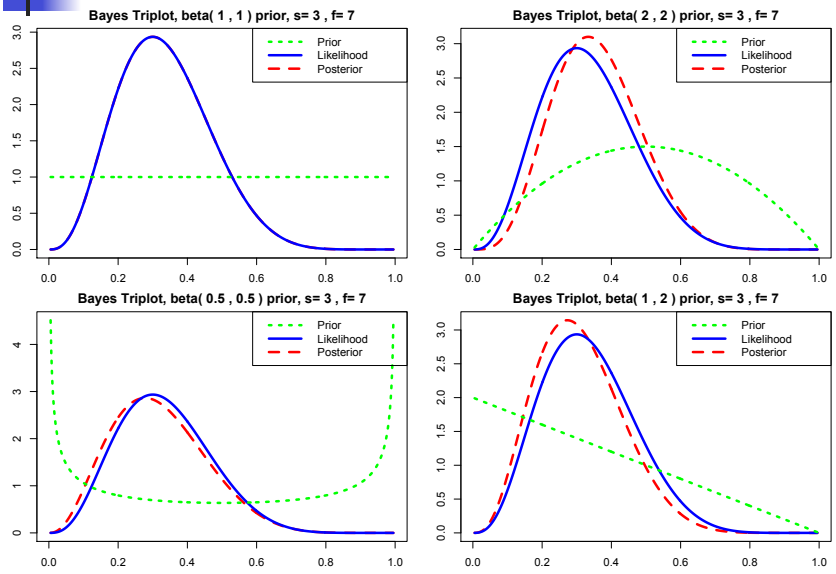


- Point estimation:
  - Mean = 0.333
  - Median = 0.324
  - Mode = 0.300
- Interval estimation:
  - Equal tail 95% credible interval: [0.109, 0.610]
  - 95% HPD: [0.101,0.581]

Interpretation: there is a 95% probability that the test sensitivity lies between [0.101, 0.581]

[Note: we obtain probability statements about θ]

Introduction to Bayesian Computation: conjugate models



Overview of the Bayesian approach

- Hypothesis testing:
  - Hypotheses: H_0 vs. H_1 [simple vs. simple]
  - Prior probabilities: $Pr(H_0)$ & $P(H_1)$
  - Likelihood: $P(Data|H_0)$ & $P(Data|H_1)$
  - Posterior probabilities:
    $P(H_0|Data) = P(H_0)\ P(Data|H_0) / P(Data)$
    where $P(Data) = P(Data|H_0)\ P(H_0) + P(Data|H_1)\ P(H_1)$
  - Odds:

$$\frac{P(H_0 \mid Data)}{P(H_1 \mid Data)} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Posterior Odds = Likelihood Ratio x Prior Odds
(a.k.a. Bayes Factor)

Overview of the Bayesian approach

- Strength of evidence provided by Bayes Factor

BF will partially eliminate the influence of the prior and emphasizes the role of data

Bayes Factor	Evidence in favor of H_0 versus H_1
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 32	Strong
32 to 100	Very strong
>100	Decisive

Overview of the Bayesian approach

Back to example:

- Test results among 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)

Introduction to Bayesian Computation: conjugate models

Back to example:

- Test results among 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)

H_0: θ = 0.5 versus H_1: θ ≠ 0.5 Priors for hypotheses: • $P(H_0)=P(H_1)=0.5$ • Under alternative: θ ~ Beta(1,1)	> pbetat(p0=0.5, prob=0.5, ab=c(1,1), data=c(3,7)) $bf [1] 1.289063 $post [1] 0.5631399

- The posterior probability of the null hypothesis is 0.56

Overview of the Bayesian approach

- Prediction:
  - Prior predictive distribution:

$$P(Y) = \int P(Y \mid \theta) P(\theta) d\theta$$

  - Posterior Predictive Distribution of Y_{NEW}

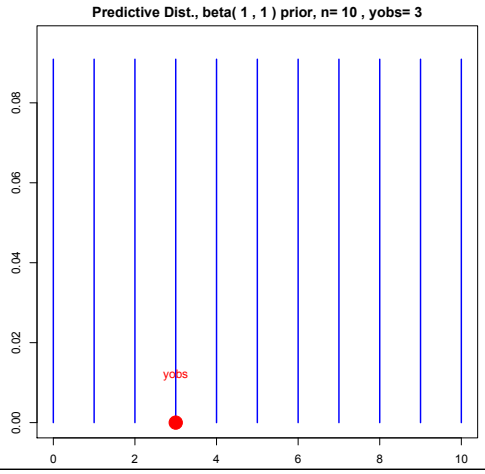
$$P(Y_{NEW} \mid Data) = \int P(Y_{NEW} \mid Data, \theta) P(\theta \mid Data) d\theta$$
$$= \int P(Y_{NEW} \mid \theta) P(\theta \mid Data) d\theta$$

  - Uses:
    - Design and (predictive) power calculations
    - Sequential monitoring
    - Model checking
    - Decision making
    - …

Introduction to Bayesian Computation: conjugate models

Prior Predictive Distribution

```
> pbetap(ab=c(1,1), n=10, s=0:10)
 [1] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909
 [7] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909

> predplot(prior=c(1,1), n=10, yobs=3)
```



Posterior Predictive Distribution

```
> pbetap(ab=c(4,8), n=1, s=0:1)
[1] 0.6666667 0.3333333
```


CLINICAL TRIALS ARTICLE

Clinical Trials 2014; **11**: 485–493

The utility of Bayesian predictive probabilities for interim monitoring of clinical trials

Benjamin R Saville[a], Jason T Connor[b,c], Gregory D Ayers[a] and JoAnn Alvarez[a]

Background Bayesian predictive probabilities can be used for interim monitoring of clinical trials to estimate the probability of observing a statistically significant treatment effect if the trial were to continue to its predefined maximum sample size.

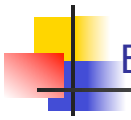
Purpose We explore settings in which Bayesian predictive probabilities are advantageous for interim monitoring compared to Bayesian posterior probabilities, *p*-values, conditional power, or group sequential methods.

Results For interim analyses that address prediction hypotheses, such as futility monitoring and efficacy monitoring with lagged outcomes, only predictive probabilities properly account for the amount of data remaining to be observed in a clinical trial and have the flexibility to incorporate additional information *via* auxiliary variables.

Limitations Computational burdens limit the feasibility of predictive probabilities in many clinical trial settings. The specification of prior distributions brings additional challenges for regulatory approval.

Conclusions The use of Bayesian predictive probabilities enables the choice of logical interim stopping rules that closely align with the clinical decision-making process. *Clinical Trials* 2014; **11**: 485–493. http://ctj.sagepub.com

CASE STUDY

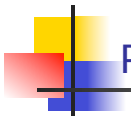


Background

- Interim analyses for stopping/continuing trials are one form of adaptive trials
- Various metrics for decisions of stopping
  - Frequentist: Multi-stage, group sequential designs, conditional power
  - Bayesian: Posterior distributions, predictive power, Bayes factors
- Question: Why and when should we use Bayesian predictive probabilities for interim monitoring?

Why interim analyses?

- Questions they can address:
  - Is there convincing evidence in favor of the null or alternative hypotheses?
    - Evidence presently shown by data
  - Is the trial likely to show convincing evidence in favor of the alternative hypothesis if additional data are collected?
    - Prediction of what evidence will be available later
- Important factors to consider:
  - ethical imperative to avoid treating patients with ineffective or inferior therapies
  - inefficient allocation of resources



Predictive Probability of Success

- Definition:
  - The probability of achieving a successful (signicant) result at a future analysis, given the current interim data
- Computation:
  - Obtained by integrating the data likelihood over the posterior distribution (i.e. we integrate over future possible responses) and predicting the future outcome of the trial
- Decision making:
  - Efficacy rules based either on Bayesian posterior distributions (fully Bayesian) or frequentist p-values (mixed Bayesian-frequentist)

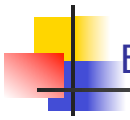
Computation via Simulation

1) At an interim analysis, sample the parameter of interest from the current posterior given current data.

2) Complete the dataset by sampling future samples, observations not yet observed at the interim analysis, from the predictive distribution.

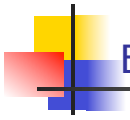
3) Use the complete dataset to calculate success criteria (p-value, posterior probability). If success criteria is met (e.g. p-value < 0.05), the trial is a success.

4) Repeat steps 1-3 a total of B times; the predictive probability (PPoS) is the proportion of simulated trials that achieve success.



Example

- Trial:
  - Single arm Phase II study of 100 patients measuring binary outcome (favorable response to treatment)
  - Goal: compare proportion to a gold standard 50% response rate
- Model: $X \sim Bin(p;N = 100)$ where
  - p = probability of response in the study population
  - N = total number of patients
- Prior: $p \sim Uniform(0,1) = Beta(1,1)$



Example

- Trial Design:
  - Trial is a success if the posterior probability that the proportion exceeds the gold standard is greater than $\eta=0.95$, that is,

$$Pr(p > 0.5|x) > \eta$$

  - Success if 59 or more of 100 patients respond
    - $Pr(p > 0.50|x = 58; n = 100) = 0.944$
    - $Pr(p > 0.50|x = 59; n = 100) = 0.963$
  - 3 interim analyses monitoring at 20, 50, and 75 patients

CASE STUDY

Table 2. Definitions of key measures and methods for illustrative example

Measure/method	Description	Formula
p-value	Probability of observing a proportion equal to or greater than x/n given $H_0 : p = p_0$	$\sum_{i=x}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$
Posterior probability	Bayesian posterior probability that proportion exceeds the null value p_0	$\Pr(p>p_0 \mid x) = \int_{p_0}^{1} f(x\mid p)\pi(p)/f(x)\,dp$
Predictive probability	Bayesian predictive probability of statistical significance at N given x/n and $\pi(p)$	$\sum_{y=0}^{m}[I\{\Pr(p>p_0 \mid x,y,N)>\eta\}f(y\mid x)]$
Conditional power	Frequentist probability of statistical significance at N given x/n and assumed p_0^*	$\sum_{y=0}^{m}\left[I\left\{\sum_{i=x+y}^{N}\binom{N}{i}p_0^i(1-p_0)^{N-i}<\alpha\right\}f(y\mid p^*)\right]$
Repeated testing of H_1	Method of monitoring for futility based on *p*-value for test of alternative hypothesis	$p\text{-value} = \sum_{i=0}^{x}\binom{n}{i}p_1^i(1-p_1)^{n-i}$
Group sequential	Frequentist design for interim monitoring that allocates Type I/II errors across interim analyses	Varies by method
Stochastic curtailment	Method that estimates the probability of statistical significance at some future sample size	Varies by method

n and N: number of patients at interim and final sample sizes, respectively; $m = N - n$: number of remaining patients yet to be observed in the study; x: number of successes observed at the interim analysis; y: number of successes yet to be observed in the remaining patients; p_0 and p_1: proportion of successes under the null hypothesis and altemative hypotheses; p^*: estimated or assumed value of p required for conditional power computation; α and η: criteria required to demonstrate 'statistical significance' for p-value or posterior probability, respectively, $I()$: indicator function taking the value 1 if expression is true and 0 if otherwise; $\pi(p)$: beta (1, 1) = 1: prior distribution of p, uniform over (0,1); $f(x) = \int_0^1 f(x\mid p)\pi(p)\,dp$: marginal likelihood or normalizing constant; $f(y\mid x) = \int_0^1 f(y\mid p)f(p\mid x)dp = \int_0^1 f(y\mid p)f(x\mid p)\pi(p)/f(x)\,dp$ = beta-binomial$(m, 1+x, 1+n-x)$: *Bayesian posterior predictive distribution of* y given x;
$f(x\mid p) = \binom{n}{x}p^x(1-p)^{n-x}$: data likelihood of x given p for n patients observed by interim;
$f(y\mid p) = \binom{m}{y}p^y(1-p)^{m-y}$: data likelihood of y given p for remaining m patients.

CASE STUDY

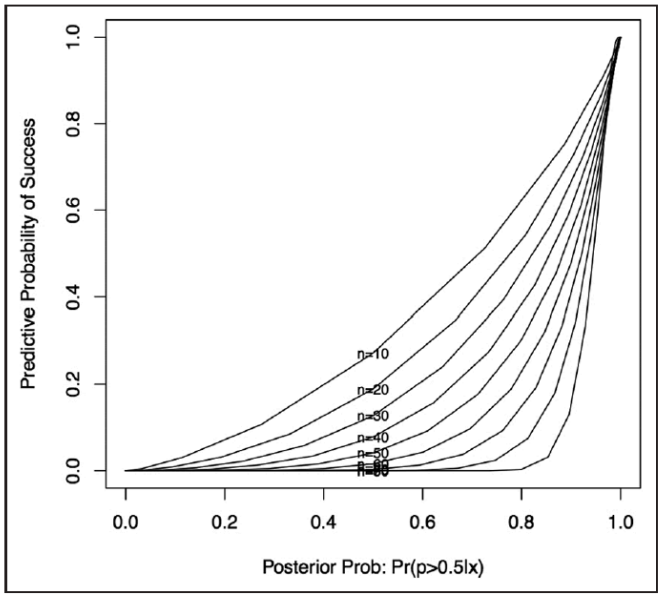
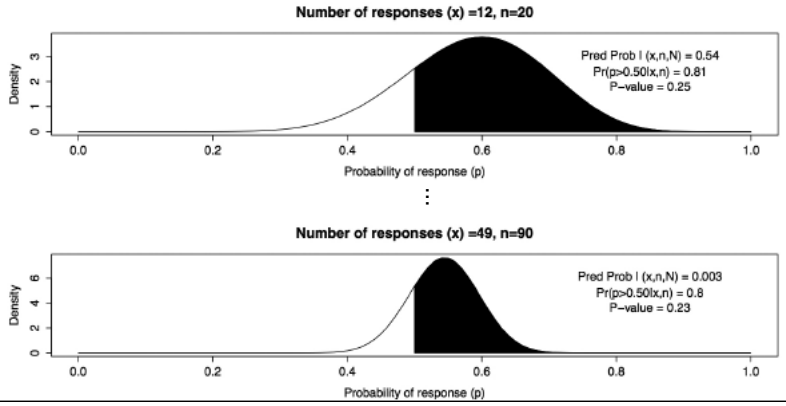


Figure 4. Predictive probability of success versus posterior estimate $\Pr(p>0.50|x)$ by interim sample size n, with maximum sample size $N = 100$ and posterior threshold $\eta = 0.95$.

Table 1. Illustrative example

n_j	x_j	m_j	y_j^*	p-value	Pr(p>0.5)	CP_{H_a}	CP_{MLE}	PP
20	12	80	47	0.25	0.81	0.90	0.64	0.54
50	28	50	31	0.24	0.80	0.73	0.24	0.30
75	41	25	18	0.24	0.79	0.31	0.060	0.086
90	49	10	10	0.23	0.80	0.013	0.002	0.003

n_j and x_j: the number of patients and successes at interim analysis j; MLE: maximum likelihood estimate; m_j : number of remaining patients at interim analysis j; y_j^*: minimum number of successes required to achieve success; CP_{H_a} and CP_{MLE}: conditional power based on original H_a or MLE; PP: Bayesian predictive probability of success.



R function to compute PP

```
PP <- function(n.total= 100, nullp = 0.5, eta=0.95, data=c(12,8), prior.par=c(1,1), B=1000){
  # posterior
  post.par <- data + prior.par

  # samples from posterior distribution
  post.sample <- rbeta(B, post.par[1], post.par[2])

  # samples new values of x (extending to the maximum sample size)
  x.new <- rbinom(B, size=n.total-sum(data), post.sample)

  # organize data with first column number of 'responses' and second 'non responses'
  data.new <- cbind(x.new, n.total-sum(data)-x.new)

  # posterior parameters given predicted data
  post.pred.par <- cbind(data.new[,1] + post.par[1], data.new[,2]+ post.par[2])

  # posterior probability that P(p > nullp |data)
  post.pred <- pbeta(nullp, post.pred.par[,1], post.pred.par[,2], lower.tail=FALSE)

  # posterior predictive probability of success
  PP <- mean(post.pred > eta)
  return(PP)
}
```

```
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(12,20-12), prior.par=c(1,1), B=1000)
[1] 0.55
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(28,50-28), prior.par=c(1,1), B=1000)
[1] 0.307
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(41,75-41), prior.par=c(1,1), B=1000)
[1] 0.081
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(49,90-49), prior.par=c(1,1), B=1000)
[1] 0.003
```

Comparison of approaches to inference

Comparison of approaches to inference

- MLE
- Intervals based on values of θ with large likelihood
- Evidence against null hypothesis via p-values

Traditional Approach

Estimation/Testing satisfying long-run properties (repeated sampling)
- Unbiased estimation
- Confidence intervals
- Type I/II error rates

- Minimax
- Admissibility…

TABLE 4 *A taxonomy of six possible 'philosophical' approaches to health technology assessment, depending on their objective and their quantitative use of prior information*

	Objective		
	Inference (estimation)	**Hypothesis testing**	**Decision (loss function)**
No prior	Fisherian	Neyman–Pearson	Classical decision theory
Prior	Proper Bayesian	'Bayes's factors'	Full decision-theoretic Bayesian

Comparison of approaches to inference

- Sequential Analysis
  - Data periodically analyzed and study stops if there are sufficiently convincing results
  - Traditional Approach:
    - Identifies "stopping boundaries" with fixed overall Type I error and chooses designs with minimum type II error for particular alternative hypotheses
    - At the end of the study, p-values and confidence intervals are adjusted for the sequential nature of the design
  - Bayesian Approach:
    - Posterior distribution following each observation becomes the prior for the next
    - Posterior distribution does not depend on the stated stopping procedure (data influence the posterior only through the likelihood)

Comparison of approaches to inference

- Sequential use of Bayes Theorem:

$$1.\ p(\theta \mid y_1) \propto p(\theta)p(y_1 \mid \theta)$$

$$\begin{aligned} 2.\ p(\theta \mid y_1, y_2) &\propto p(\theta)p(y_1, y_2 \mid \theta) \\ &\propto p(\theta)p(y_1 \mid \theta)p(y_2 \mid \theta) \\ &\propto p(\theta \mid y_1)p(y_2 \mid \theta) \end{aligned}$$

  - Posterior distribution using initial prior p(θ) given all the data is the same to that obtained sequentially where posterior for the current observation becomes the prior for the next observation.

Comparison of approaches to inference

- P-values and Bayes factors (BF)
  - Example:
    - Model
      - Y ~ Binomial(n, θ)
    - Parameter
      - θ=True unknown population proportion of preference for A
    - Hypotheses
      - H_0: θ = 0.5 versus H_1: θ ≠ 0.5
      - Under alternative θ ~ U(0,1)=Beta(1,1)

$$P(Y = y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{(n-y)}$$

Recall:

$$\frac{P(H_0 \mid Data)}{P(H_1 \mid Data)} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Bayes Factor

Comparison of approaches to inference

- Bayes Factor (BF):

$$P(Data \mid H_0) = P(Y = y \mid \theta = 0.5) = \binom{n}{y} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{n-y} = \binom{n}{y} 2^{-n}$$

$$P(Data \mid H_1) = P(Y = y \mid \theta \neq 0.5) = \int P(Y = y \mid \theta) p(\theta) d\theta = \ldots = \frac{1}{n+1}$$

$$BF = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} = \binom{n}{y} \frac{n+1}{2^n}$$

- Alternative: Likelihood-based Bayes Factor (Minimum BF)

$$P(Data \mid H_1) = P(Y = y \mid \theta = \hat{\theta}_{MLE}) = \binom{n}{y} \left(\frac{y}{n}\right)^y \left(1 - \frac{y}{n}\right)^{n-y}$$

$$BF_{\min} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} = \frac{\frac{1}{2^n}}{\left(\frac{y}{n}\right)^y \left(1 - \frac{y}{n}\right)^{n-y}}$$

Comparison of approaches to inference

Sample Size	Preference for A	Estimate	P-value (One-sided)	Min. BF	BF
20	15	0.750	0.02	0.07	0.31
200	115	0.575	0.02	0.10	1.20
2000	1046	0.523	0.02	0.12	4.30
2000000	1001445	0.500	0.02	0.12	139.8

- Interpretation of p-values is dependent on sample size!
- Minimum BFs obey the Likelihood Principle, but have similar qualitative behavior to P-values
- Proper BFs can, for large samples relative to the prior precision, support the null hypothesis when a classical analysis would lead to its rejection.
  - This is known as the Lindley's paradox
    - Explanation: For large sample sizes, a p-value can be small even if the data support parameter values very close to the null hypothesis. Such data may be unlikely under the null, but even more unlikely under the alternative that spreads the prior over a wide range of values. Thus, the BF can support the null when the significance test would reject it.

Large Sample Properties

$$y = (y_1, \ldots, y_n) \quad \text{where} \quad y_i \sim p(y_i \mid \theta) \quad \text{and} \quad p(y|\theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$$

$$\text{Let: } I(\theta) = E\left[-\frac{\partial^2 \log P(Y \mid \theta)}{\partial \theta_i \partial \theta_j} \right] \text{ (Fisher information)}$$

Large Sample Properties

- Likelihood-based Inference (MLE)

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\hat{\theta})\right)$$

- Bayesian Inference

$$\theta \sim N\left(\hat{\theta}, I^{-1}(\hat{\theta})\right)$$

  - Thus, the posterior distribution will give essentially the same asymptotic estimates and intervals as the maximum likelihood estimator. However, note that the posterior distribution is a distribution of θ given $\hat{\theta}$ whereas the previous result gives the sampling distribution of $\hat{\theta}$ given θ.

Large Sample Properties

To convince you of the previous result, suppose the parameter is uni-dimensional. Note that we get the same density functions:

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\hat{\theta})\right)$$

$$p(\hat{\theta} \mid \theta) = \frac{1}{\sqrt{2\pi I^{-1}(\hat{\theta})}} \exp\left[-\frac{1}{2I^{-1}(\hat{\theta})}(\hat{\theta}-\theta)^2\right]$$

$$\theta \sim N\left(\hat{\theta}, I^{-1}(\hat{\theta})\right)$$

$$p(\theta \mid \hat{\theta}) = \frac{1}{\sqrt{2\pi I^{-1}(\hat{\theta})}} \exp\left[-\frac{1}{2I^{-1}(\hat{\theta})}(\theta-\hat{\theta})^2\right]$$

Bayesian GLM

Generalized Linear Regression Models

- Mean: $E[Y_i \mid X_{i1}, X_{i2}, \ldots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where g is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
  - Linear regression model
    $$g(\mu_i) = \mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Logistic regression model
    $$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Probit regression model
    $$g(\mu_i) = \Phi^{-1}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Poisson regression model
    $$g(\mu_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$

Bayesian GLM

- Mean: $E[Y_i \mid X_{i1},\ldots,X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where g is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
- **Priors**:
  - Regression parameters: $(\beta_0, \beta_1, \beta_2, \ldots \beta_p)$
  - "Nuisance" parameters (e.g. in linear regression σ^2)
- Note:
  - Regression coefficients have the same interpretation (e.g. difference in means; log-odds ratio; etc)
  - Interpretation of inferential results are different (e.g. posterior mean; probability that the regression parameter lies in some interval; etc)

Bayesian GLM in R

- We will use the following packages
  - INLA
  - arm
- Different approaches to estimation of GLMs
  - Approximate posterior inference (Bayesian CLT)
- Advantages:
  - Syntax very similar to those we reviewed for traditional GLMs
  - No need for heavy programming (e.g. MCMC methods)
- Disadvantages:
  - Approximate method under small samples
  - Constrained by model formulations handled by the packages

Bayesian GLM in R: INLA package

- Integrated Nested Laplace Approximations (INLA)
  - Alternative to MCMC in (latent) Gaussian models
  - Regression Model:

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \sum_{k=1}^{q} f_k(\tilde{X}_{ik}) + \varepsilon_i$$

$f_k(.)$: unknown functions of covariates $\tilde{X}$

β_j : linear effects of covariates X

ε_i : unstructured terms

  - Assumption in latent Gaussian models:

Gaussian Prior for: $\beta_0, \{\beta_j\}, \{f_j(.)\}, \{\varepsilon_i\}$

Bayesian GLM in R: INLA package

- Latent model:

Let z represent the collection of all Gaussian variables:

$$\beta_0, \{\beta_j\}, \{f_j(.)\}, \{\varepsilon_i\}$$

with distribution $\pi(z \mid \theta_1)$ with mean 0 precision matrix $Q(\theta_1)$.

- Model: $\pi(y \mid z, \theta_2)$
- Prior: Let $\theta = (\theta_1, \theta_2)$ with prior $\pi(\theta)$.
- Via Gaussian & Laplace approximations:

$$\tilde{\pi}(\theta \mid y) \propto \left. \frac{\pi(z, \theta, y)}{\tilde{\pi}_G(z \mid \theta, y)} \right|_{z = z^*(\theta)}$$

$z^*(\theta)$: mode of $\pi(z|\theta, y)$

$\tilde{\pi}_G$: Gaussian approximation of $\pi(z|\theta, y)$

Latent Models in INLA

Model	Name	Description
Independent random variables	iid	indep.pdf
Linear	linear	linear.pdf
Constrained linear	clinear	clinear.pdf
Random walk of order 1	rw1	rw1.pdf
Random walk of order 2	rw2	rw2.pdf
Continuous random walk of order 2	crw2	crw2.pdf
Model for seasonal variation	seasonal	seasonal.pdf
Model for spatial effect	besag	besag.pdf
Model for spatial effect	besagproper	besagproper.pdf
Model for weighted spatial effects	besag2	besag2.pdf
Model for spatial effect + random effect	bym	bym.pdf
Autoregressive model of order 1	ar1	ar1.pdf
Autoregressive model of order p	ar	ar.pdf
The Ornstein-Uhlenbeck process	ou	ou.pdf
User defined structure matrix, type 0	generic0	generic0.pdf
User defined structure matrix, type1	generic1	generic1.pdf
User defined structure matrix, type2	generic2	generic2.pdf
Model for correlated effects with Wishart prior (dimension 1, 2, 3, 4 and 5).	iid1d, iid2d, iid3d, iid4d, iid5d	iid123d.pdf
Classical random effect model	z	z.pdf
Random walk of 2nd order on a lattice	rw2d	rw2d.pdf
Gaussian field with Matern covariance function	matern2d	matern2d.pdf
Classical measurement error model	mec	mec.pdf
Berkson measurement error model	meb	meb.pdf
Spatial lag model	slm	slm.pdf
Sigmodial and reverse sigmodial	sigm, revsigm	sigm.pdf

Likelihoods

Negative Binomial	nbinomial	nbinomial.pdf
Poisson	poisson	poisson.pdf
Binomial	binomial	binomial.pdf
CBinomial	cbinomial	cbinomial.pdf
Gaussian	gaussian	gaussian.pdf
Skew Normal	sn	sn.pdf
Student-t	T	Student-t.pdf
Gaussian model for stochastic volatility	stochvol	stochvolgaussian.pdf
Student-t model for stochastic volatility	stochvol.t	stochvolt.pdf
NIG model for stochastic volatility	stochvol.nig	stochvolnig.pdf
Zero inflated Poisson	zeroinflated.poisson.0 zeroinflated.poisson.1 zeroinflated.poisson.2	zeroinflated.pdf
Zero inflated Binomial	zeroinflated.binomial.0 zeroinflated.binomial.1	zeroinflated.pdf
Zero inflated negative Binomial	zeroinflated.nbinomial.0 zeroinflated.nbinomial.1 zeroinflated.nbinomial.2	zeroinflated.pdf
Zero inflated beta binomial (type 0/1)	zeroinflated.betabinomial.0 zeroinflated.betabinomial.1	zeroinflated.pdf
Zero inflated beta binomial (type 2)	zeroinflated.betabinomial.2	zeroinflatedbetabin.pdf
Generalised extreme value distribution (GEV)	gev	gev.pdf
Beta	beta	beta.pdf
Gamma	gamma	gamma.pdf
Beta-Binomial	betabinomial	betabinomial.pdf
Logistic distribution	logistic	logistic.pdf
Exponential (Survival models)	exponential	exponential.pdf
Weibull (Survival model)	weibull	weibull.pdf
LogLogistic (Survival model)	loglogistic	loglogistic.pdf
LogNormal (Survival model)	lognormal	lognormal.pdf
Cox model (Survival model)	coxph	coxph.pdf

Priors on hyperparameters

Model	Name	Description
Normal distribution	`normal` `gaussian`	gaussian.pdf
Log-gamma distribution	`loggamma`	prior-loggamma.pdf
Improper flat prior	`flat`	prior-flat.pdf
Truncated Normal distribution	`logtnormal` `logtgaussian`	log-tnormal.pdf
Improper flat prior on the log scale	`logflat`	various-flat.pdf
Improper flat prior on the 1/ log scale	`logiflat`	various-flat.pdf
Wishart prior	`wishart`	iid123d.pdf
Beta for correlations	`betacorrelation`	betacorrelation.pdf
Logit of a Beta	`logitbeta`	logitbeta.pdf
Define your own prior	`expression:`	expression.pdf
Define your own prior	`table:`	table.pdf

Bayesian GLM in R: arm package

- Builds on a modification of glm()
  - Uses priors on an augmented regression
  - Uses an approximate EM algorithm to update regression coefficients
    - Gelman, Jakulin, Grazia, Pittau, Su, 2008. A Weekly Informative Default Prior Distribution for Logistic and Other Regression Models. The Annals of Applied Statistics, 2,1360-1383.

Bayesian GLM in R: arm package

- Augmentation Idea (context linear models):

Matrix Formulation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

$$\text{In short}: Y = X\beta + \varepsilon$$

Prior: $\beta_j \sim N(m_j, v_j^2),\ j = 0,\ldots,p$

Augmented Data: $Y^* = \begin{bmatrix} Y \\ m \end{bmatrix}, X^* = \begin{bmatrix} X \\ I_p \end{bmatrix}$

Bayesian GLM in R: arm package

bayesglm {arm} **R Documentation**

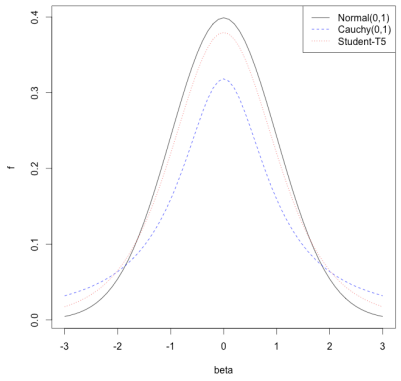
Bayesian generalized linear models.

Description

Bayesian functions for generalized linear modeling with independent normal, t, or Cauchy prior distribution for the coefficients.

Usage

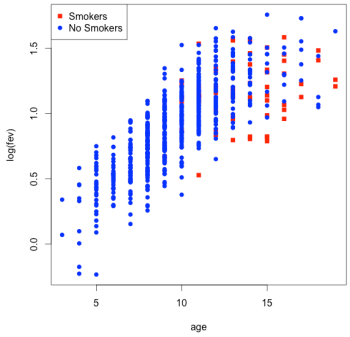
```
bayesglm (formula, family = gaussian, data,
   weights, subset, na.action,
   start = NULL, etastart, mustart,
   offset, control = glm.control(...),
   model = TRUE, method = "glm.fit",
   x = FALSE, y = TRUE, contrasts = NULL,
   drop.unused.levels = TRUE,
   prior.mean = 0,
   prior.scale = NULL,
   prior.df = 1,
   prior.mean.for.intercept = 0,
   prior.scale.for.intercept = NULL,
   prior.df.for.intercept = 1,
   min.prior.scale=1e-12,
   scaled = TRUE, keep.order=TRUE,
   drop.baseline=TRUE, n.iter = 100,
   print.unnormalized.log.posterior=FALSE,
   Warning=TRUE,...)
```



Multiple Linear Regression in R

- FEV dataset: (located on the class web pages) contains data on 654 children.
  - seqnbr — case number (the numbers 1 to 654)
  - subjid — subject identification number (unique for each different child)
  - age — subject age at time of measurement (years)
  - fev — measured FEV (liters per second)
  - height — subject height at time of measurement (inches)
  - sex — subject sex (1 = male, 2 = female)
  - smoke — smoking habits (1 = yes, 2 = no)

- Our goal is to assess the association between FEV and smoking status adjusting for age.



Multiple Linear Regression in R

```
## read FEV data set
data = read.table("data/fev.txt", col.names=c("seqnbr", "subjid", "age", "fev", "height", "sex", "smoke"))

## examine a few entries of the data set
head(data)

 seqnbr subjid age   fev height sex smoke
1      1    301   9 1.708   57.0   2     2
2      2    451   8 1.724   67.5   2     2
3      3    501   7 1.720   54.5   2     2
4      4    642   9 1.558   53.0   1     2
5      5    901   9 1.895   57.0   1     2
6      6   1701   8 2.336   61.0   2     2

## summarize the variables
summary(data)

 seqnbr          subjid           age              fev             height           sex            smoke
 Min.   :  1.0   Min.   :  201   Min.   : 3.000   Min.   :0.791   Min.   :46.00   Min.   :1.000   Min.   :1.000
 1st Qu.:164.2   1st Qu.:15811   1st Qu.: 8.000   1st Qu.:1.981   1st Qu.:57.00   1st Qu.:1.000   1st Qu.:2.000
 Median :327.5   Median :36071   Median :10.000   Median :2.547   Median :61.50   Median :1.000   Median :2.000
 Mean   :327.5   Mean   :37170   Mean   : 9.931   Mean   :2.637   Mean   :61.14   Mean   :1.486   Mean   :1.901
 3rd Qu.:490.8   3rd Qu.:53638   3rd Qu.:12.000   3rd Qu.:3.119   3rd Qu.:65.50   3rd Qu.:2.000   3rd Qu.:2.000
 Max.   :654.0   Max.   :90001   Max.   :19.000   Max.   :5.793   Max.   :74.00   Max.   :2.000   Max.   :2.000


## scatter plot of log(fev) by age
plot(log(fev) ~ age, data=data)


## scatter plot of log(fev) by age, but stratified by smoking status
plot(log(fev) ~ age, type="n", data=data)
points(log(fev) ~ age, col='red', pch=15, data=data[data$smoke==1,])
points(log(fev) ~ age, col='blue', pch=16, data=data[data$smoke==2,])
legend("topleft", c("Smokers", "No Smokers"), col=c("red", "blue"), pch=c(15,16))
```

Bayesian GLM in R: arm package

```
> ## -- Normal priors for regression coefficients (with mean=0 and scale=10)
> fit <- bayesglm(log(fev) ~ smoke + age, data=data, family=gaussian, prior.mean=0,
   prior.scale=10, prior.df=Inf)
> summary(fit3)

Call:
bayesglm(formula = log(fev) ~ smoke + age, family = gaussian,
    data = data, prior.mean = 0, prior.scale = 10, prior.df = Inf)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.71124  -0.13458   0.00104   0.14910   0.60260

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.156900   0.075106  -2.089  0.03709 *
smoke        0.089923   0.030049   2.993  0.00287 **
age          0.090767   0.003046  29.801  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.044219911)

    Null deviance: 72.526  on 653  degrees of freedom
Residual deviance: 28.920  on 654  degrees of freedom
AIC: -175.58

Number of Fisher Scoring iterations: 5
```

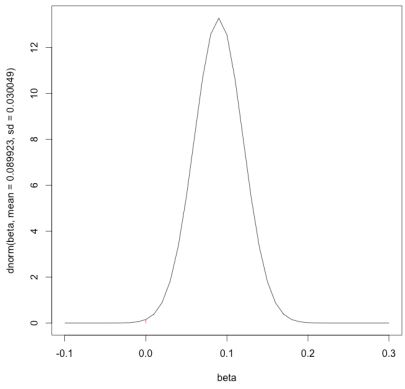
This can be interpreted as posterior mean/median & posterior standard deviations of the regression coefficients

This can be interpreted as two-sided posterior tail probabilities of "no effect"…

Bayesian GLM in R: arm package

- More formally, the posterior probabilities are:

$$2 \times \min(P(\beta_j \leq 0 \mid data), P(\beta_j \geq 0 \mid data))$$



Bayesian GLM in R: arm package

```
Call:
lm(formula = log(fev) ~ smoke + age, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.71124 -0.13458  0.00104  0.14909  0.60261

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.156915   0.075279  -2.084  0.03751 *
smoke        0.089927   0.030118   2.986  0.00293 **
age          0.090768   0.003053  29.733  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2108 on 651 degrees of freedom
Multiple R-squared:  0.6012,  Adjusted R-squared:  0.6
F-statistic: 490.8 on 2 and 651 DF,  p-value: < 2.2e-16
```

Traditional inference

Bayesian inference

```
Call:
bayesglm(formula = log(fev) ~ smoke + age, family = gaussian,
    data = data, prior.mean = 0, prior.scale = 10, prior.df = Inf)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-0.71124  -0.13458   0.00104   0.14910   0.60260

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.156900   0.075106  -2.089  0.03709 *
smoke        0.089923   0.030049   2.993  0.00287 **
age          0.090767   0.003046  29.801  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.04421991)

    Null deviance: 72.526  on 653  degrees of freedom
Residual deviance: 28.920  on 654  degrees of freedom
AIC: -175.58
```

Exercise:

Draw similarities & differences (what explains similarities?)

Bayesian GLM in R: INLA package

Bayesian linear regression: FEV data

```
> library(INLA)

> fit = inla(log(fev)~ smoke + age, data=data)

> fit$summary.fix
                  mean          sd  0.025quant    0.5quant    0.975quant        mode          kld
(Intercept) -0.15691453 0.075213737 -0.30461595 -0.15691665 -0.009339742 -0.15691453 4.170371e-14
smoke        0.08992702 0.030091877  0.03083392  0.08992617  0.148969457  0.08992702 4.070901e-14
age          0.09076807 0.003050123  0.08477837  0.09076798  0.096752630  0.09076807 4.661605e-14

> fit$summary.hy
                                        mean       sd 0.025quant 0.5quant 0.975quant     mode
Precision for the Gaussian observations 22.58597 1.250476   20.20882 22.55905   25.12505 22.51306
```

Bayesian GLM in R: INLA package

Bayesian linear regression: FEV data

```
fit.prior1 = inla(log(fev)~ smoke + age, data=data,
          control.family = list(
            hyper = list(
              prec = list(
                prior = "normal",
                param = c(0, 10)
              )
            )
           )
)
fit.prior1$summary.fix
fit.prior1$summary.hy
```

Making prior assumptions explicit

```
> fit.prior1$summary.fix
                mean          sd  0.025quant    0.5quant    0.975quant        mode          kld
(Intercept) -0.15691451 0.079087993 -0.31222632 -0.15691674 -0.001736104 -0.15691451 4.714742e-14
smoke        0.08992701 0.031641908  0.02778911  0.08992612  0.152011540  0.08992701 4.721406e-14
age          0.09076807 0.003207235  0.08446975  0.09076798  0.097060977  0.09076807 4.216079e-14
> fit.prior1$summary.hy
                                              mean       sd 0.025quant 0.5quant 0.975quant     mode
Precision for the Gaussian observations 20.42931 1.169583   18.20924   20.403   22.80734 20.35767
```


Survival Models: Notation

- Let T be a continuous non-negative random variable representing survival times of individuals in some population
  - Density function (pdf): f(t)
  - Distribution function (cdf): F(t)
    - Fraction of people dying by time t
  - Survival function: S(t)
    - Fraction of people surviving at time t
  - Hazard function: h(t)
    - Instantaneous risk of death
  - Cumulative Hazard: H(t)

Survival Models: Relationships

$$h(t) = \frac{f(t)}{S(t)}$$

$$H(t) = \int_0^t h(u)\,du$$

$$F(t) = \int_0^t f(u)\,du$$

$$S(t) = 1 - F(t) = \exp(-H(t))$$

$$f(t) = h(t)S(t) = h(t)\exp(-H(t))$$

- Likelihood contribution for a subject who dies

$$f(t) = h(t)S(t)$$

- Likelihood contribution for a subject who is censored

$$S(t)$$

- Thus, if d is the indicator of death, we can write:

$$[h(t)]^d\, S(t)$$

Survival Models: Proportional Hazards

- Proportional Hazards (PH) Model:

$$h(t) = h_0(t)\exp\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right)$$

- Parametric vs Semi-parametric PH model?
  - What is the form of the baseline hazard ($h_0(.)$) function?

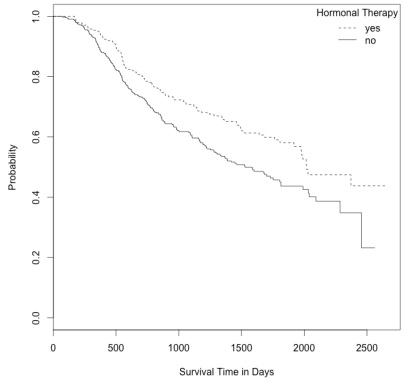
PH regression models in R

- Data from the German Breast Cancer Study Group 2 contains the observations of 686 women:

horTh	hormonal therapy, a factor at two levels no and yes.
age	age of the patients in years.
menostat	menopausal status, a factor at two levels pre(premenopausal) and post (postmenopausal)
tsize	tumor size (in mm)
tgrade	tumor grade, a ordered factor at levels I < II < III.
pnodes	number of positive nodes
progrec	progesterone receptor (in fmol)
estrec	estrogen receptor (in fmol)
time	recurrence free survival time (in days)
cens	censoring indicator (0- censored, 1- event).

PH regression models in R

```
> ## -- data publicly available in an R-package
> data("GBSG2", package="TH.data")
> summary(GBSG2)
 horTh          age          menostat      tsize          tgrade       pnodes          progrec          estrec            time
 no :440   Min.   :21.00   Pre :290   Min.   :  3.00   I  : 81   Min.   : 1.00   Min.   :   0.0   Min.   :   0.00   Min.   :   8.0
 yes:246   1st Qu.:46.00   Post:396   1st Qu.: 20.00   II :444   1st Qu.: 1.00   1st Qu.:   7.0   1st Qu.:   8.00   1st Qu.: 567.8
           Median :53.00              Median : 25.00   III:161   Median : 3.00   Median :  32.5   Median :  36.00   Median :1084.0
           Mean   :53.05              Mean   : 29.33             Mean   : 5.01   Mean   : 110.0   Mean   :  96.25   Mean   :1124.5
           3rd Qu.:61.00              3rd Qu.: 35.00             3rd Qu.: 7.00   3rd Qu.: 131.8   3rd Qu.: 114.00   3rd Qu.:1684.8
           Max.   :80.00              Max.   :120.00             Max.   :51.00   Max.   :2380.0   Max.   :1144.00   Max.   :2659.0
      cens
 Min.   :0.0000
 1st Qu.:0.0000
 Median :0.0000
 Mean   :0.4359
 3rd Qu.:1.0000
 Max.   :1.0000
```



PH regression models in R

```
> ## (Semi-Parametric) Cox PH model
> fit1 <- coxph(Surv(time, cens) ~ horTh, data=GBSG2)
> summary(fit1)
Call:
coxph(formula = Surv(time, cens) ~ horTh, data = GBSG2)

  n= 686, number of events= 299

            coef exp(coef) se(coef)      z Pr(>|z|)
horThyes -0.3640    0.6949   0.1250 -2.911   0.0036 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         exp(coef) exp(-coef) lower .95 upper .95
horThyes    0.6949      1.439    0.5438    0.8879

Concordance= 0.543  (se = 0.015 )
Rsquare= 0.013   (max possible= 0.995 )
Likelihood ratio test= 8.82  on 1 df,   p=0.002977
Wald test            = 8.47  on 1 df,   p=0.003602
Score (logrank) test = 8.57  on 1 df,   p=0.003425
```

```
> ## Parametric survival (Weibull regression)
> library(eha)
> fit3 <- weibreg(Surv(time, cens) ~ horTh, data=GBSG2)
> summary(fit3)
Call:
weibreg(formula = Surv(time, cens) ~ horTh, data = GBSG2)

Covariate           Mean       Coef Exp(Coef)  se(Coef)    Wald p
horTh
             no     0.604      0         1           (reference)
            yes     0.396     -0.393     0.675     0.125     0.002

log(scale)                     7.608  2015.149     0.058     0.000
log(shape)                     0.251     1.285     0.050     0.000

Events                    299
Total time at risk        771400
Max. log. likelihood      -2632.1
LR test statistic         10.4
Degrees of freedom        1
Overall p-value           0.00128307
```


Bayesian PH regression models in R: Non-parametric

```
> library(INLA)
> ## Bayesian non-parametric PH model
> fit <- inla(inla.surv(time, cens) ~ horTh, family="coxph",data=GBSG2)
> summary(fit)

Call:

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0948          0.5117          0.0574          0.6639

Fixed effects:
                mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept) -7.7078 0.1403    -7.9948  -7.7039    -7.4426 -7.6965   0
horThyes    -0.3660 0.1249    -0.6145  -0.3650    -0.1237 -0.3628   0

Random effects:
Name            Model
 baseline.hazard   RW1 model

Model hyperparameters:
                                  mean    sd      0.025quant 0.5quant 0.975quant mode
Precision for baseline.hazard 1451.61  943.88  363.52    1221.72  3902.23     849.37

Expected number of effective parameters(std dev): 9.484(1.091)
Number of equivalent replicates : 497.48

Marginal Likelihood:  -1379.80
```

Bayesian PH regression models in R:
Parametric

```
> ## Bayesian parametric PH model
> fit <- inla(inla.surv(time, cens) ~ horTh, family="weibull",data=GBSG2)
> summary(fit)

Call:
c("inla(formula = inla.surv(time, cens) ~ horTh, family = \"weibull\", ",  "    data = GBSG2)")

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0698          1.2193          0.0485          1.3376

Fixed effects:
               mean     sd 0.025quant 0.5quant 0.975quant    mode   kld
(Intercept) -9.5518 0.4442   -10.2282  -9.3908    -8.9047 -9.3598 1e-04
horThyes    -0.3891 0.1248    -0.6373  -0.3880    -0.1470 -0.3859 0e+00

The model has no random effects

Model hyperparameters:
                                  mean   sd     0.025quant 0.5quant 0.975quant mode
alpha parameter for weibull 1.2651 0.0749 1.1438     1.2557   1.4339     1.2290

Expected number of effective parameters(std dev): 2.005(0.00)
Number of equivalent replicates : 342.15

Marginal Likelihood:  -2641.95
```


LONGITUDINAL ANALYSIS OF SERIAL MEASUREMENTS OF FREE AND TOTAL PSA AMONG MEN WITH AND WITHOUT PROSTATIC CANCER

JAY D. PEARSON, ALBERT A. LUDERER, E. JEFFREY METTER, ALAN W. PARTIN, DANIEL W. CHAN, JAMES L. FOZARD, AND H. BALLENTINE CARTER

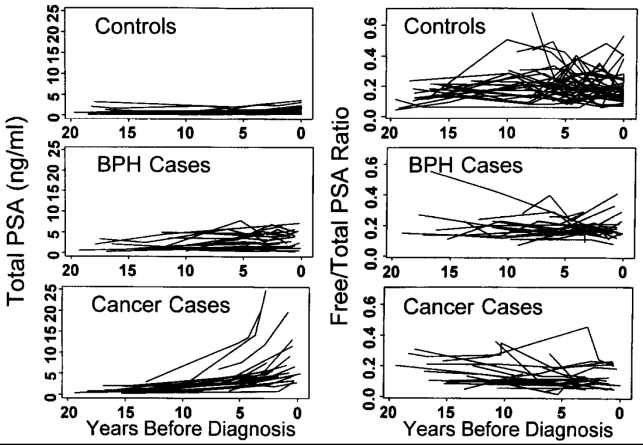
Example: Longitudinal data

Urology 48(6A):4-9,1996

ABSTRACT

Objectives. Evaluation of free and total serum prostate specific antigen (PSA) levels before diagnosis of prostate cancer.

Methods. Free and total PSA levels were measured on frozen sera samples of 26 men with no history of prostate disease (controls), 29 men with a histologic diagnosis of benign prostatic hyperplasia (BPH) made at simple prostatectomy (BPH cases), and 23 men with a histologic diagnosis of prostatic cancer (cancer cases). Longitudinal regression analysis was used to evaluate PSA levels as a function of years before diagnosis of prostate disease.



Covariation in the socioeconomic determinants of self rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA

S V Subramanian, Daniel Kim, Ichiro Kawachi

J Epidemiol Community Health 2005;**59**:664–669. doi: 10.1136/jech.2004.025742

See end of article for authors' affiliations

Correspondence to: Dr S V Subramanian, Department of Society, Human Development and Health, Harvard School of Public Health, 677 Huntington Avenue, 7th Floor, Boston, MA 02115, USA; svsubram@hsph.harvard.edu

Accepted for publication 22 October 2004

Objective: To investigate individual level determinants of self rated health and happiness, as well as the extent of community level covariation in health and happiness.
Design: Multivariate multilevel regression analysis of self rated poor health and unhappiness at level 1, nested within 24 118 people at level 2, nested within 36 communities at level 3. Data were obtained from the 2000 social capital benchmark survey.
Setting: USA communities.
Participants: 24 118 adults.
Main outcome measures: Self reported fair/poor health; and a single item measure of subjective wellbeing.
Results: Controlling for demographic markers, a strong income and education gradient was seen for self rated poor health and unhappiness, with the gradient being stronger for poor health. Community level correlations between self rated poor health and happiness were stronger (0.65) than the individual level correlations (0.16) between the two outcomes.
Conclusion: Poor health and unhappiness are highly positively correlated within individuals, and communities that are healthier tend to be happier and vice versa.

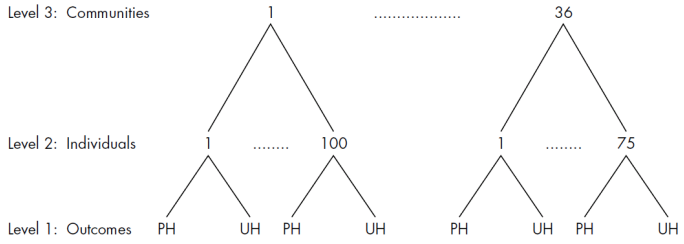


Figure 1 Multivariate multilevel structure of responses (PH, poor health; UH, unhappy) at level 1 nested within individuals at level 2 nested within communities at level 3

American Journal of Epidemiology
Vol. 145, No. 6
Printed in U.S.A.

Spatial Analysis of the Distribution of Lyme Disease in Wisconsin

Uriel Kitron[1] and James J. Kazmierczak[2]

Surveillance measures for human cases of Lyme disease in Wisconsin were compared and associated with tick distribution and vegetation coverage. During 1991–1994, 1,759 confirmed human cases of Lyme disease reported to the Wisconsin Division of Health were assigned a county of residence, but only 329 (19%) could be assigned with certainty a county of exposure. Distributions of cases by county of exposure and residence were often consistent from year to year. Tick distribution in 46 of 72 Wisconsin counties was mapped based on collections by researchers, statewide surveys of infested deer, and submissions from the public. Satellite data were used to calculate a normalized difference vegetation index (NDVI) for each county. A geographic information system (GIS) was used to map distributions of human Lyme disease cases, ticks, and degree of vegetation cover. Human case distribution by county of exposure was significantly correlated with tick distribution; both were positively correlated with high NDVI values in spring and fall, when wooded vegetation could be distinguished from agricultural crops in the satellite image. Statistical analysis of spatial patterns using a measure of spatial autocorrelation indicated that counties with most human cases and ticks were clustered in parts of western Wisconsin. A map delineating the counties with highest risk for Lyme disease transmission was generated based on numbers of exposed human cases and tick concentrations. *Am J Epidemiol* 1997;145:558–66.

geographic information systems; geography; Lyme disease; remote sensing; spatial analysis; ticks

Example: Spatial data

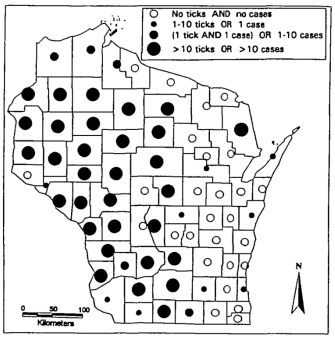


FIGURE 4. Lyme disease endemicity in Wisconsin counties, 1970–1995, as determined by county of exposure for human cases and known distribution of *Ixodes scapularis*.

Original Article

Time Series Analysis of Incidence Data of Influenza in Japan

Ayako Sumi[1], Ken-ichi Kamo[2], Norio Ohtomo[3], Keiji Mise[4], and Nobumichi Kobayashi[1]

[1]Department of Hygiene, Sapporo Medical University School of Medicine, Sapporo, Japan
[2]Department of Liberal Arts and Sciences, Sapporo Medical University, Sapporo, Japan
[3]Natural Energy Research Center Co. Ltd., Sapporo, Japan
[4]Center for Medical Education, Department of Admissions, Sapporo Medical University, Sapporo, Japan

Received September 29, 2009; accepted July 8, 2010; released online November 13, 2010

ABSTRACT

Background: Much effort has been expended on interpreting the mechanism of influenza epidemics, so as to better predict them. In addition to the obvious annual cycle of influenza epidemics, longer-term incidence patterns are present. These so-called interepidemic periods have long been a focus of epidemiology. However, there has been less investigation of the interepidemic period of influenza epidemics. In the present study, we used spectral analysis of influenza morbidity records to indentify the interepidemic period of influenza epidemics in Japan.

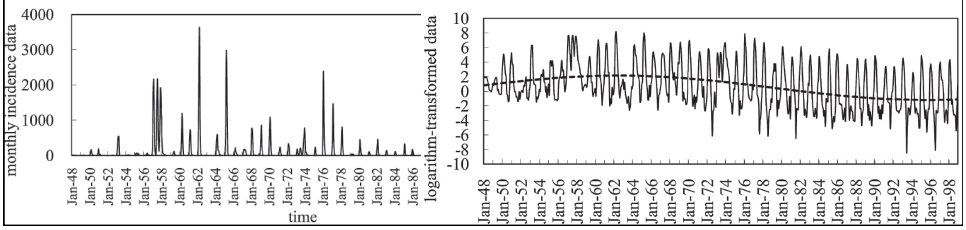
Methods: We used time series data of the monthly incidence of influenza in Japan from January 1948 through December 1998. To evaluate the incidence data, we conducted maximum entropy method (MEM) spectral analysis, which is useful in investigating the periodicities of shorter time series, such as that of the incidence data used in the present study. We also conducted a segment time series analysis and obtained a 3-dimensional spectral array.

Results: Based on the results of power spectral density (PSD) obtained from MEM spectral analysis, we identified 3 periodic modes as the interepidemic periods of the incidence data. Segment time series analysis revealed that the amount of amplitude of the interepidemic periods increased during the occurrence of influenza pandemics and decreased when vaccine programs were introduced.

Conclusions: The findings suggest that the temporal behavior of the interepidemic periods of influenza epidemics is correlated with the magnitude of cross-reactive immune responses.

Example: Time series data

a.

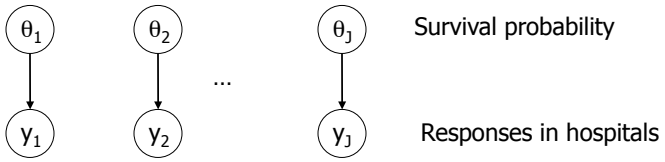


Modeling of Correlated Data: Motivation

- Degree of "similarity" may help with prediction!
  - Lyme disease incidence rates more similar in closer neighborhoods
  - Incidence rates of flu more similar within "short" time periods
  - Incidence rates of flu with similar seasonal patterns (e.g. Winter) across years
  - Happiness rates more similar from individuals within the same communities
  - …

Hierarchical Model Example:

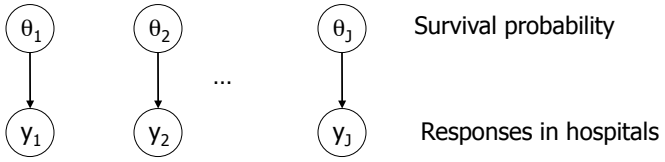
- Goal:
  - Study the effectiveness of cardiac treatments



Independent Data
(Separate analysis using data from each study)

Hierarchical Model Example:

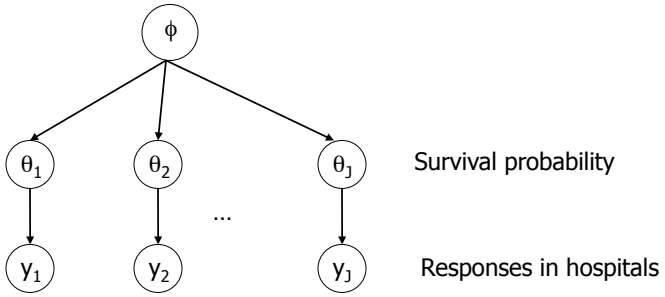
- Goal:
  - Study the effectiveness of cardiac treatments



It may be reasonable to expect that estimates of θ_j's, which represent a sample of hospitals, should be related to each other: $\theta_j \sim \pi(\phi)$, j=1,...,J.

Hierarchical Model Example:

- Goal:
  - Study the effectiveness of cardiac treatments



This implies, marginally, correlation between observations!

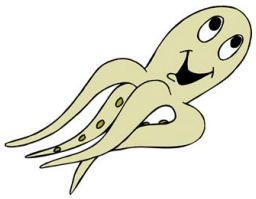
Hierarchical Model Example:

- Goal:
  - Study the effectiveness of cardiac treatments
    - θ_j: survival probability for patients in hospital j
    - ϕ : overall survival probability
  - Inference:
    - Estimate θ_j's borrowing strength of information from all other hospitals
    - Estimate ϕ taking into account the variability among hospitals

Hierarchical Model:
Exchangeability

- Definition: $Y_1, \ldots, Y_n$ are judged **exchangeable** if the probability $P(Y_1, \ldots, Y_n)$ is unaffected by permutations of the labels attached to the variables.
  - Example:

$$\text{If } P(Y_1,Y_2,Y_3)=P(Y_2,Y_1,Y_3)=P(Y_2,Y_3,Y_1)=$$
$$=P(Y_1,Y_3,Y_2)=P(Y_3,Y_1,Y_2)=$$
$$=P(Y_3,Y_2,Y_1)$$



we would judge Y_1,Y_2,Y_3 exchangeable!

Hierarchical Model:
Exchangeability

- Note:
  - An infinite sequence of random variables $Y_1, Y_2, \ldots$ is exchangeable if any finite subsequence is exchangeable.
  - Independence implies exchangeability, but not conversely! That is, independence is a stronger assumption than exchangeability.

Hierarchical Model:
Exchangeability

$P(Y_1 = 0; Y_2 = 1) = P(Y_1 = 1; Y_2 = 0) = 0.09$

	0	1	Total
0	0.01	0.09	0.10
1	0.09	0.81	0.90
Total	0.10	0.90	1.00

If two random variables Y_1 and Y_2 are independent then they are exchangeable, but exchangeability does not imply independence…

$P(Y_1 = 0; Y_2 = 1) = P(Y_1 = 1; Y_2 = 0) = 0.05$

	0	1	Total
0	0.05	0.05	0.10
1	0.05	0.85	0.90
Total	0.10	0.90	1.00

Hierarchical Model:
Exchangeability

- Checking exchangeability could be difficult if we had to assess the probabilities of all permutations
- We can bypass this with a nice result…

Hierarchical Model:
Exchangeability: De Finetti's theorem

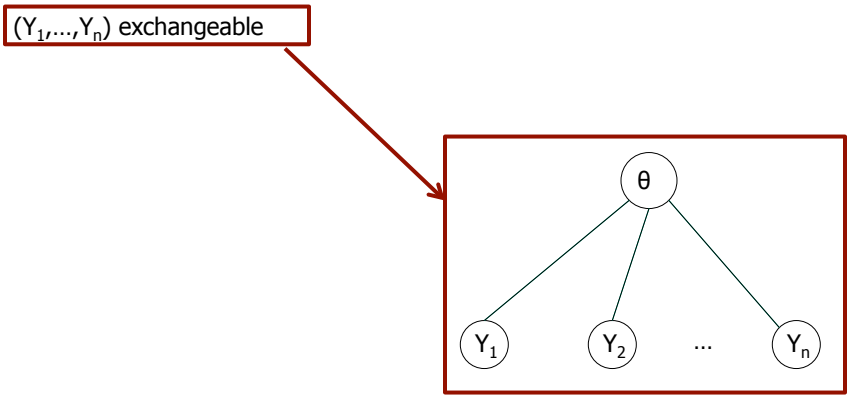
- For all infinite sequences of exchangeable random binary variables $\{Y_1, Y_2, \ldots\}$, there corresponds a distribution function F on (0,1) such that for all n and $k \leq n$,

$$P[(k, n-k)] = \int_0^1 \theta^k (1-\theta)^{n-k} \, dF(\theta)$$

- What is "cool" about this?
  - Justifies the Bayesian approach:
    - If one is willing to assume that a collection of 0-1 variables is exchangeable, then one is prepared to re-phrase the model into a sampling Bernoulli model with success probability θ that is itself random with probability distribution F (the prior).
    - The theorem does not tell us anything about what the distribution F should be!

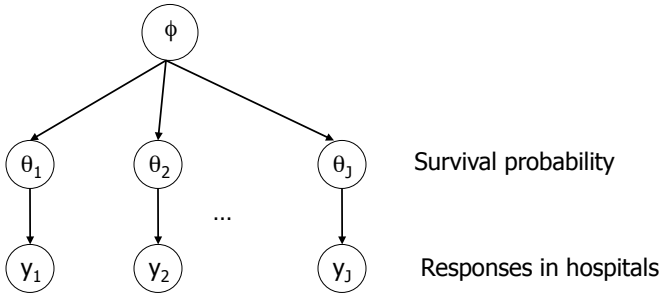
Hierarchical Model:
Exchangeability: De Finetti's theorem

Representation:



Hierarchical Model:
Exercise (back to example)

- Where would we assume exchangeability?



Hierarchical Models

- Definition:
  - A Bayesian Hierarchical model is a statistical model ($f(x|\theta)$, $\pi(\theta)$) where the prior distribution $\pi(\theta)$ is decomposed in conditional distributions

    $$\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \ldots, \pi_n(\theta_{n-1}|\theta_n)$$

    and a marginal distribution $\pi_{n+1}(\theta_n)$ such that

    $$\pi(\theta)=\int\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \ldots, \pi_n(\theta_{n-1}|\theta_n)\ \pi_{n+1}(\theta_n)\ d\theta_1 \ldots\ d\theta_n$$

    Parameters θ_i are called <u>hyperparameters</u> of level I

- Higher level of hierarchy assumes known hyperparameters.
  - Difficult to check propriety of posteriors with improper priors
    - Proper distributions which are almost vague can also approach impropriety with undesirable modeling results
  - <u>Sensitivity analysis is very important in hierarchical modeling</u>

Hierarchical Models

- Approach to building complex models by specifying a series of conditional distributions
- Parameters in the model can be regarded as related or connected in some way by the structure of the problem
- Typically data have multi-level/hierarchical structure (observational units grouped into larger units)
  - Example: students are grouped into classes, which are grouped into schools, which are grouped by districts…
- Levels of inference dependent on scientific questions of interest
  - Example: Multi-center clinical trial
    - Magnitude of an "average" treatment effect?
    - Magnitude of treatment effect in each center?
    - Amount of variation of the effect across centers?
    - …

Efficacy of BCG Vaccine in the Prevention of Tuberculosis

Meta-analysis of the Published Literature

Graham A. Colditz, MD, DrPH; Timothy F. Brewer, MD, MPH; Catherine S. Berkey, DSc, MA; Mary E. Wilson, MD; Elisabeth Burdick, MSc; Harvey V. Fineberg, MD, PhD; Frederick Mosteller, PhD

Objective.—To quantify the efficacy of BCG vaccine against tuberculosis (TB).

Data Sources.—MEDLINE with index terms *BCG vaccine*, *tuberculosis*, and *human*. Experts from the Centers for Disease Control and Prevention and the World Health Organization, among others, provided lists of all known studies.

Study Selection.—A total of 1264 articles or abstracts were reviewed for details on BCG vaccination, concurrent vaccinated and unvaccinated groups, and TB outcome; 70 articles were reviewed in depth for method of vaccine allocation used to create comparable groups, equal surveillance and follow-up for recipient and concurrent control groups, and outcome measures of TB cases and/or deaths. Fourteen prospective trials and 12 case-control studies were included in the analysis.

Data Extraction.—We recorded study design, age range of study population, number of patients enrolled, efficacy of vaccine, and items to assess the potential for bias in study design and diagnosis. At least two readers independently extracted data and evaluated validity.

Data Synthesis.—The relative risk (RR) or odds ratio (OR) of TB provided the measure of vaccine efficacy that we analyzed. The protective effect was then computed by 1−RR or 1−OR. A random-effects model estimated a weighted average RR or OR from those provided by the trials or case-control studies. In the trials, the RR of TB was 0.49 (95% confidence interval [CI], 0.34 to 0.70) for vaccine recipients compared with nonrecipients (protective effect of 51%). In the case-control studies, the OR for TB was 0.50 (95% CI, 0.39 to 0.64), or a 50% protective effect. Seven trials reporting tuberculous deaths showed a protective effect from BCG vaccine of 71% (RR, 0.29; 95% CI, 0.16 to 0.53), and five studies reporting on meningitis showed a protective effect from BCG vaccine of 64% (OR, 0.36; 95% CI, 0.18 to 0.70). Geographic latitude of the study site and study validity score explained 66% of the heterogeneity among trials in a random-effects regression model.

Conclusion.—On average, BCG vaccine significantly reduces the risk of TB by 50%. Protection is observed across many populations, study designs, and forms of TB. Age at vaccination did not enhance predictiveness of BCG efficacy. Protection against tuberculous death, meningitis, and disseminated disease is higher than for total TB cases, although this result may reflect reduced error in disease classification rather than greater BCG efficacy.

(*JAMA*. 1994;271:698-702)

CASE STUDY

Table 1.—Reports From Clinical Trials Providing Estimates of Efficacy of BCG Vaccine Against Cases of Tuberculosis (TB) and TB Death That Were Used in the Meta-analysis*

Source, y	Population		Cases of TB			TB Death		
	BCG	No BCG	BCG	No BCG	RR	BCG	No BCG	RR
Aronson,[23] 1948†	123	139	4	11	0.41	0	4	0.14
Ferguson and Simes,[40] 1949	306	303	6	29	0.20	2	9	0.22
Rosenthal et al,[42] 1960‡	231	220	3	11	0.26	0	4	0.12
Hart and Sutherland,[41] 1977	13 598	12 867	62	248	0.24	...	...	...
Frimodt-Moller et al,[45] 1973	5069	5808	33	47	0.80	...	...	...
Stein and Aronson,[44] 1953	1541	1451	180	372	0.46	...	...	...
Vandiviere et al,[43] 1973	2545	629	8	10	0.20	...	...	...
Madras,[15] 1980§	88 391	88 391	505	499	1.01	...	...	...
Coetzee and Berjak,[39] 1968‖	7499	7277	29	45	0.63	...	...	...
Rosenthal et al,[46] 1961¶	1716	1665	17	65	0.25	1	6	0.16
Comstock et al,[47] 1974	50 634	27 338	186	141	0.71	8	12	0.36
Comstock and Webster,[48] 1969#	2498	2341	5	3	1.56	...	...	...
Comstock et al,[49] 1976#	16 913	17 854	27	29	0.98	...	...	...
Aronson et al,[51] 1958**	1541	1451	...	...	...	13	68	0.18
Levine and Sackett,[50] 1948††	566	528	...	...	...	8	8	0.93
Overall RR (95% confidence interval)				0.49 (0.34-0.70)			0.29 (0.16-0.53)	

*RR indicates relative risk. Ellipses indicate data not reported.
†Infants study.
‡TB households.
§Data based on 7.5-year follow-up of entire population. We estimated the population numbers because they were not reported.
‖Miners randomized during year 3 of the trial had a truncated follow-up period; we used person-years of follow-up to estimate total sample size.
¶Non-TB households.
#Follow-up sample sizes were not reported. We assumed follow-up was comparable in BCG and no BCG groups.
**This report on deaths is based on the same trial as Stein and Aronson, 1953.[44]
††Data after 1932 recruitment.

CASE STUDY

Systematic Reviews and Meta-Analysis

- Motivation:
  - Many individual clinical trials are not large enough to answer the questions of interest reliably
- Solutions
  - Advocacy for large trials
    - Not always feasible
  - Informal evidence synthesis from different studies
    - Possibility of biased selection of evidence
  - Formal systematic review

Systematic Reviews and Meta-Analysis

- Goals of Systematic Reviews:
  - To review systematically the available evidence from a particular research area
  - To provide quantitative summaries of the results from each study
  - To combine the results across studies if appropriate; such combination of results leads to greater statistical power in estimating treatment effects
  - To asses the amount of variability between studies
  - To estimate the degree of benefit associated with a particular study treatment
  - To identify study characteristics associated with particularly effective treatments.

Systematic Reviews and Meta-Analysis

- Components of Systematic Reviews:
  - Qualitative:
    - Description of available trials in terms of relevance and methodological strengths and weaknesses
  - Quantitative
    - Means of combining results from different studies
    - This is known as Meta-Analysis
- Critical Step:
  - Study selection

Systematic Reviews and Meta-Analysis

- Statistical Methodology
  - Fixed effects models
    - Each individual study used to estimate a common, unknown, overall pooled effect
  - Random effects models
    - Each individual study has its own underlying effect, which in turn are used to estimated a common population effect.
    - Accounts for two sources of heterogeneity:
      - Within-study heterogeneity
      - Between-study heterogeneity

Systematic Reviews and Meta-Analysis

- Fixed-Effects (Mantel-Haenszel):

$$\text{Pooled Effect}: \bar{Y} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \quad \text{with} \quad Var(\bar{Y}) = \frac{1}{\sum_{i=1}^{k} W_i}$$

k : number of studies

Y_i : effect size in the i-th study

W_i : weight (inverse of within-study variance for i-th study)

Systematic Reviews and Meta-Analysis

- Random-Effects (DerSimonian-Laird):

$$Y_i = \mu_i + \sigma_i \varepsilon_i \quad for \ i = 1,...,k$$

$$\mu_i \sim N(\mu, \tau^2); \ \varepsilon_i \sim N(0,1)$$

$$\text{Pooled Effect}: \bar{Y} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \quad ; \quad \text{Weights}: W_i = \frac{1}{V_i^2 + \hat{\tau}^2}$$

$$\hat{\tau}^2 = \begin{cases} 0, \text{ if } Q < k-1 \\ (Q - k + 1)/U, \text{ if } Q > k-1 \end{cases}$$

$$Q = \sum_{i=1}^{k} W_i (Y_i - \bar{Y})^2; \quad U = (k-1)(\bar{W} - s_w^2 / kW)$$

Systematic Reviews and Meta-Analysis

- Heterogeneity is very likely in meta-analysis
  - Many possible sources of heterogeneity
  - Estimating how these various factors affect the effect size is often of considerable interest and importance
    - Meta-regression!

BCG Example

Efficacy of BCG Vaccine
in the Prevention of Tuberculosis

Meta-analysis of the Published Literature

Graham A. Colditz, MD, DrPH; Timothy F. Brewer, MD, MPH; Catherine S. Berkey, DSc, MA; Mary E. Wilson, MD; Elisabeth Burdick, MSc; Harvey V. Fineberg, MD, PhD; Frederick Mosteller, PhD

- Bacille Calmette Guerin (BCG)
  - Most widely used vaccine against tuberculosis (TBC)
- Expanded Data: publicly available in R
  - 13 clinical trials of BCG investigating efficacy in the treatment of tuberculosis
    - Number of subjects with TB with our without BCG vaccination
  - Heterogeneity among trials may be explained by geographic location and year
- Efficacy measure: Odds Ratio (OR)

CASE STUDY

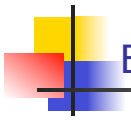
BCG Example

- Data:

trial	author	year	tpos	tneg	cpos	cneg	ablat	alloc
1	Aronson	1948	4	119	11	128	44	random
2	Ferguson & Simes	1949	6	300	29	274	55	random
3	Rosenthal et al	1960	3	228	11	209	42	random
4	Hart & Sutherland	1977	62	13536	248	12619	52	random
5	Frimodt-Moller et al	1973	33	5036	47	5761	13	alternate
6	Stein & Aronson	1953	180	1361	372	1079	44	alternate
7	Vandiviere et al	1973	8	2537	10	619	19	random
8	TPT Madras	1980	505	87886	499	87892	13	random
9	Coetzee & Berjak	1968	29	7470	45	7232	27	random
10	Rosenthal et al	1961	17	1699	65	1600	42	systematic
11	Comstock et al	1974	186	50448	141	27197	18	systematic
12	Comstock & Webster	1969	5	2493	3	2338	33	systematic
13	Comstock et al	1976	27	16886	29	17825	33	systematic

- The 13 studies provide data in terms of *2x2* tables in the form:

	TB positive	TB negative
vaccinated group	tpos	tneg
control group	cpos	cneg



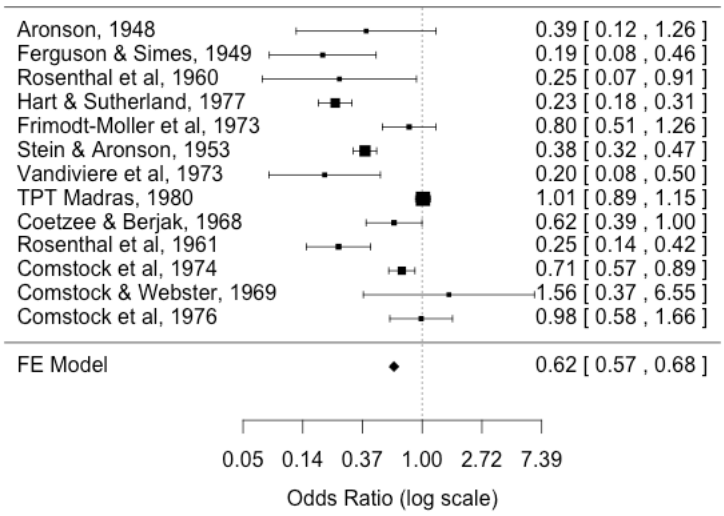
BCG Example

```
## Meta-Analysis
library(metafor)

## load data
data(dat.bcg)

## Part A: frequentist analysis
##-- meta-analysis of the log odds ratio using the Mantel-Haenszel method
res.fe <- rma.mh(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg,
slab=paste(author, year, sep=", "))
### forest plot of the observed odds ratio with summary estimate
forest(res.fe, atransf=exp, xlim=c(-7,5), ylim=c(-2.5,16))

##-- meta-analysis of the log odds ratio using a random-effects model
res.re <- rma(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg,
slab=paste(author, year, sep=", "))
### add summary estimate from the random-effects model to forest plot
addpoly(res.re, atransf=exp)
### forest plot of the observed odds ratio with summary estimate
forest(res.re, atransf=exp, xlim=c(-7,5), ylim=c(-2.5,16))
```


BCG Example (A): Standard Meta-Analysis
Mantel-Haenszel

```
> res.fe

Fixed-Effects Model (k = 13)

Test for Heterogeneity:
Q(df = 12) = 163.9426, p-val < .0001

Model Results (log scale):

estimate       se     zval     pval    ci.lb    ci.ub
 -0.4734   0.0410 -11.5444   <.0001  -0.5538  -0.3930

Model Results (OR scale):

estimate    ci.lb    ci.ub
  0.6229   0.5748   0.6750

Cochran-Mantel-Haenszel Test:    CMH = 135.6889, df = 1,  p-val < .0001
Tarone's Test for Heterogeneity: X^2 = 171.7567, df = 12, p-val < .0001
```

BCG Example (A): Standard Meta-Analysis
Mantel-Haenszel

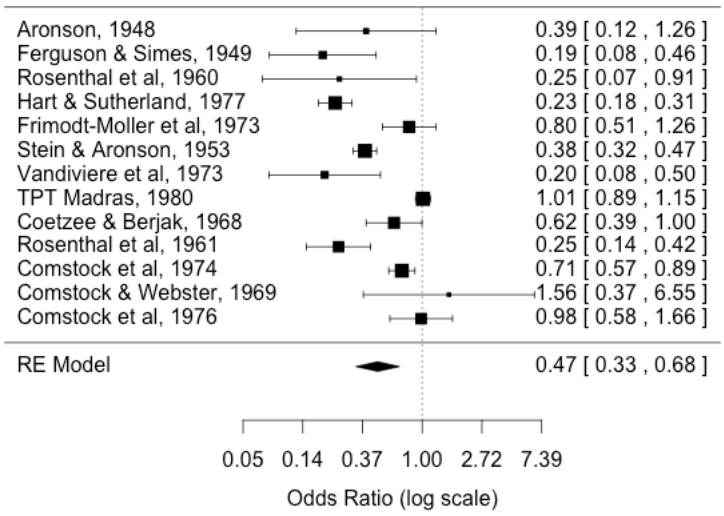


BCG Example (A): Standard Meta-Analysis
DerSimonian-Laird

```
> res.re

Random-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.3378 (SE = 0.1784)
tau (square root of estimated tau^2 value):      0.5812
I^2 (total heterogeneity / total variability):   92.07%
H^2 (total variability / sampling variability):  12.61

Test for Heterogeneity:
Q(df = 12) = 163.1649, p-val < .0001

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
 -0.7452   0.1860  -4.0057   <.0001  -1.1098  -0.3806      ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The heterogeneity test shows strong evidence of heterogeneity in the 13 trials!

BCG Example (A): Standard Meta-Analysis
DerSimonian-Laird



BCG Example

```
### meta-regression
##-- calculate log odds ratios and corresponding sampling variances
dat <- escalc(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg)
head(dat)

### random-effects model (output is the same as seen for res.re)
res <- rma(yi, vi, data=dat)
res

### average relative risk with 95% CI (this will give you the OR from combined studies)
predict(res, transf=exp)

### meta-regression model with absolute latitude and year as moderator
res.mr <- rma(yi, vi, mods = ~ ablat + year, data=dat)
res.mr
```

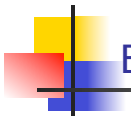
BCG Example: Meta-Regression Analysis

```
> res.mr

Mixed-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.0913 (SE = 0.0745)
tau (square root of estimated tau^2 value):             0.3022
I^2 (residual heterogeneity / unaccounted variability): 67.29%
H^2 (unaccounted variability / sampling variability):   3.06
R^2 (amount of heterogeneity accounted for):            72.96%

Test for Residual Heterogeneity:
QE(df = 10) = 25.0121, p-val = 0.0053

Test of Moderators (coefficient(s) 2,3):
QM(df = 2) = 16.2533, p-val = 0.0003

Model Results:

         estimate       se     zval    pval     ci.lb    ci.ub
intrcpt  -10.5347  27.3739  -0.3848  0.7004  -64.1865  43.1172
ablat     -0.0288   0.0095  -3.0311  0.0024   -0.0475  -0.0102  **
year       0.0055   0.0138   0.3949  0.6929   -0.0216   0.0325

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some evidence that latitude is associated with observed effect size.

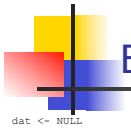
BCG Example (A): Meta-Regression Analysis

```
> res.mr1

Mixed-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.0504 (SE = 0.0449)
tau (square root of estimated tau^2 value):             0.2246
I^2 (residual heterogeneity / unaccounted variability): 57.39%
H^2 (unaccounted variability / sampling variability):   2.35
R^2 (amount of heterogeneity accounted for):            85.06%

Test for Residual Heterogeneity:
QE(df = 11) = 25.0954, p-val = 0.0088

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 25.2424, p-val < .0001

Model Results:

         estimate      se     zval    pval    ci.lb    ci.ub
intrcpt    0.3010  0.2146   1.4025  0.1608  -0.1197   0.7217
ablat     -0.0315  0.0063  -5.0242  <.0001  -0.0438  -0.0192  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



BCG Example (B): Bayesian Meta-Analysis

- We will consider several models and compare the results
- First, we need to re-organize the data…



BCG Example (B): Bayesian Meta-Analysis

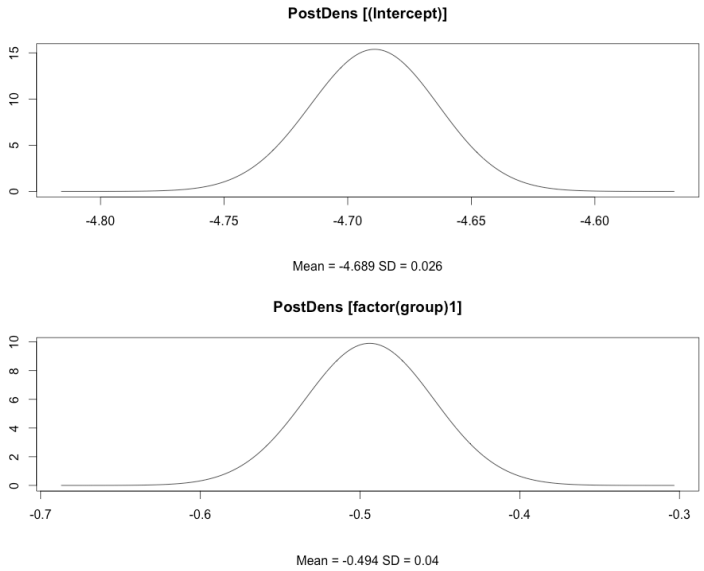
```
dat <- NULL
dat$trial <- rep(seq(1,13),2)
dat$group <- c(rep(1, 13), rep(0,13))
dat$Y <- c(dat.bcg$tpos, dat.bcg$cpos)
dat$N <- rep(NA, 26)
dat$N[1:13]  <- dat.bcg$tpos + dat.bcg$tneg
dat$N[14:26] <- dat.bcg$cpos + dat.bcg$cneg
dat$Latitude <- rep(dat.bcg$ablat,2)
dat$centeredLatitude = dat.bcg$ablat - mean(dat.bcg$ablat)
dat$Year <- rep(dat.bcg$year, 2)
dat$centeredYear = dat.bcg$year - mean(dat.bcg$year)
dat1 <- as.data.frame(dat)

$trial
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13  1  2  3  4  5  6  7  8  9 10 11 12 13

$group
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0

$Y
 [1]   4   6   3  62  33 180   8 505  29  17 186   5  27  11  29  11 248  47 372  10 499  45  65 141   3  29

$N
 [1]   123   306   231 13598  5069  1541  2545 88391  7499  1716 50634  2498 16913   139   303   220 12867  5808  1451   629 88391
[22]  7277  1665 27338  2341 17854

$Latitude
 [1] 44 55 42 52 13 44 19 13 27 42 18 33 33 44 55 42 52 13 44 19 13 27 42 18 33 33

$centeredLatitude
 [1]  10.5384615  21.5384615   8.5384615  18.5384615 -20.4615385  10.5384615 -14.4615385 -20.4615385  -6.4615385   8.5384615
[11] -15.4615385  -0.4615385  -0.4615385  10.5384615  21.5384615   8.5384615  18.5384615 -20.4615385  10.5384615 -14.4615385
[21] -20.4615385  -6.4615385   8.5384615 -15.4615385  -0.4615385  -0.4615385

$Year
 [1] 1948 1949 1960 1977 1973 1953 1973 1980 1968 1961 1974 1969 1976 1948 1949 1960 1977 1973 1953 1973 1980 1968 1961 1974 1969 1976

$centeredYear
 [1] -18.230769 -17.230769  -6.230769  10.769231   6.769231 -13.230769   6.769231  13.769231   1.769231  -5.230769   7.769231
[12]   2.769231   9.769231 -18.230769 -17.230769  -6.230769  10.769231   6.769231 -13.230769   6.769231  13.769231   1.769231
[23]  -5.230769   7.769231   2.769231   9.769231
```

Model 1 (B):

```
> fit1 = inla(Y~ factor(group), data=dat, Ntrials=N, family="binomial")
> summary(fit1)

Call:
c("inla(formula = Y ~ factor(group), family = \"binomial\", data = dat, Ntrials = N)")

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0592          0.0236          0.0308          0.1135

Fixed effects:
                  mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)    -4.6895 0.0260    -4.7408  -4.6893    -4.6388 -4.6891   0
factor(group)1 -0.4942 0.0403    -0.5734  -0.4941    -0.4152 -0.4940   0

The model has no random effects

The model has no hyperparameters

Expected number of effective parameters(std dev): 2.029(0.00)
Number of equivalent replicates : 12.81

Marginal Likelihood:  -1833.30
```

The overall posterior median OR=exp(-0.49)=0.61 (95% PCI= 0.57,0.66)
- Very similar results to those obtained using Mantel-Haenszel (fixed-effects).

Model 1 (B):



Model 2 (B):

```
> summary(fit2)

Call:
c("inla(formula = Y ~ factor(group) + centeredLatitude, family = \"binomial\", ",  "    data = dat, Ntrials = N)")

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0616          0.0279          0.0370          0.1265

Fixed effects:
                    mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)      -4.3255 0.0269    -4.3787  -4.3254    -4.2730 -4.3251   0
factor(group)1   -0.4748 0.0403    -0.5541  -0.4748    -0.3959 -0.4747   0
centeredLatitude  0.0385 0.0013     0.0360   0.0385     0.0411  0.0385   0

The model has no random effects

The model has no hyperparameters

Expected number of effective parameters(std dev): 3.028(0.00)
Number of equivalent replicates : 8.586

Marginal Likelihood:  -1452.92
```


Model 3 (B):

```
> fit3 = inla(Y~ -1 + factor(trial) + factor(group), data=dat, Ntrials=N, family="binomial")
> summary(fit3)

Call:
c("inla(formula = Y ~ -1 + factor(trial) + factor(group), family = \"binomial\", ",  "    data = dat, Ntrials = N)")

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0982          0.0437          0.0952          0.2371

Fixed effects:
                    mean     sd 0.025quant 0.5quant 0.975quant    mode kld
factor(trial)1  -2.6017 0.2667    -3.1576  -2.5905    -2.1088 -2.5674   0
factor(trial)2  -2.5823 0.1752    -2.9403  -2.5773    -2.2519 -2.5672   0
factor(trial)3  -3.2226 0.2722    -3.7921  -3.2103    -2.7214 -3.1851   0
factor(trial)4  -4.2176 0.0595    -4.3362  -4.2171    -4.1022 -4.2159   0
factor(trial)5  -4.7097 0.1132    -4.9387  -4.7074    -4.4937 -4.7027   0
factor(trial)6  -1.2581 0.0508    -1.3585  -1.2579    -1.1590 -1.2574   0
factor(trial)7  -4.8029 0.2382    -5.2989  -4.7931    -4.3622 -4.7729   0
factor(trial)8  -4.9537 0.0356    -5.0241  -4.9535    -
4.8844 -4.9532   0
factor(trial)9  -5.0772 0.1177    -5.3154  -5.0747    -4.8528 -5.0696   0
factor(trial)10 -3.4792 0.1131    -3.7075  -3.4770    -3.2630 -3.4725   0
factor(trial)11 -5.1864 0.0597    -5.3052  -5.1859    -5.0705 -5.1849   0
factor(trial)12 -6.1843 0.3541    -6.9374  -6.1635    -5.5445 -6.1201   0
factor(trial)13 -6.2250 0.1346    -6.4986  -6.2217    -5.9695 -6.2150   0
factor(group)1  -0.4784 0.0413    -0.5597  -0.4784    -0.3975 -0.4783   0

The model has no random effects

The model has no hyperparameters

Expected number of effective parameters(std dev): 14.01(0.00)
Number of equivalent replicates : 1.855

Marginal Likelihood:  -236.63
```

The overall posterior median OR=exp(-0.48)=0.62 (95% PCI= 0.57,0.67)
- Very similar results to those obtained using Mantel-Haenszel (fixed-effects).

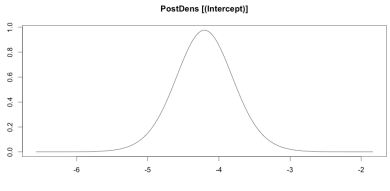
Model 4 (B):

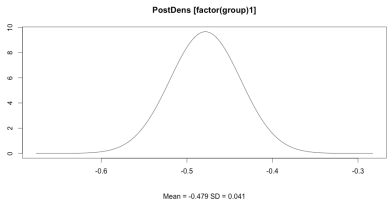
```
> fit4 = inla(Y~ factor(group) + f(trial, model="iid", param=c(0.001,0.001)), data=dat, Ntrials=N, family="binomial")
> summary(fit4)


Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0893          0.0381          0.0551          0.1825

Fixed effects:
                   mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)     -4.2043 0.4260    -5.0507  -4.2041    -3.3600 -4.2039   0
factor(group)1 -0.4785 0.0413    -0.5598  -0.4785    -0.3976 -0.4784   0

Random effects:
Name            Model
 trial    IID model

Model hyperparameters:
                      mean   sd     0.025quant 0.5quant 0.975quant mode
Precision for trial 0.4633 0.1930 0.1802      0.4335   0.9236     0.3733

Expected number of effective parameters(std dev): 13.84(0.0541)
Number of equivalent replicates : 1.879

Marginal Likelihood:  -209.55
```

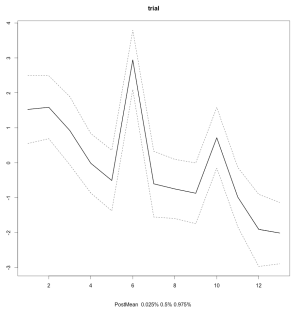
The overall posterior median OR=exp(-0.48)=0.62 (95% PCI= 0.57,0.67)
Posterior median precision = 0.43 (posterior median variance = 1/.43=2.33)

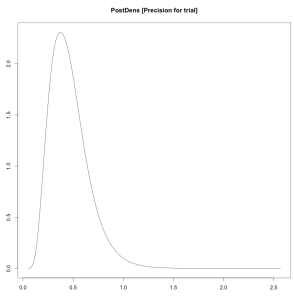
Estimated variance under frequentist is much smaller (since it doesn't account for uncertainty in random effects)

Model 4 (B):









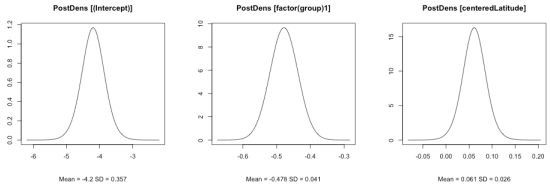
Model 5 (B):

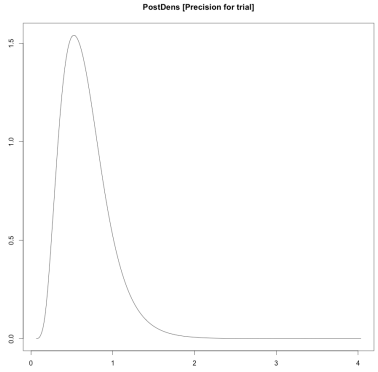
```
> summary(fit5)

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0856          0.0441          0.0602          0.1900

Fixed effects:
                    mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)      -4.2001 0.3572    -4.9112  -4.1997    -3.4922 -4.1991   0
factor(group)1   -0.4782 0.0413    -0.5595  -0.4782    -0.3973 -0.4781   0
centeredLatitude  0.0612 0.0256     0.0103   0.0612     0.1121  0.0612   0

Random effects:
Name         Model
 trial   IID model

Model hyperparameters:
                      mean   sd     0.025quant 0.5quant 0.975quant mode
Precision for trial 0.6697 0.2941 0.2467     0.6219   1.3772     0.5245

Expected number of effective parameters(std dev): 13.78(0.0743)
Number of equivalent replicates : 1.887

Marginal Likelihood:  -214.00
```

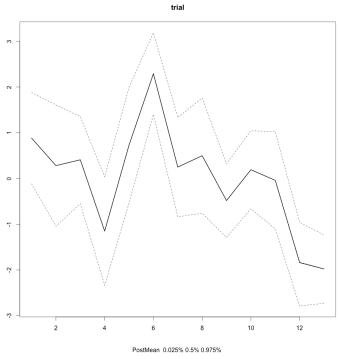
The overall posterior median OR=exp(-0.48)=0.62 (95% PCI= 0.57,0.67)
Posterior median precision = 0.62 (posterior median variance = 1/0.62=1.61)

Improved inference about precision [heterogeneity partially explained by Latitude]

Model 5 (B):





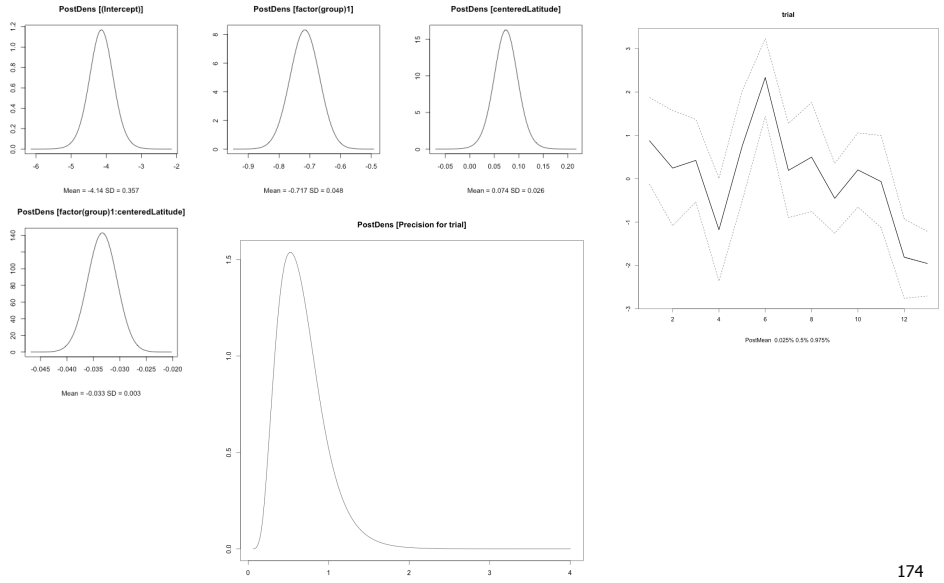


Model 6 (C):

```
> summary(fit6)

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0933          0.0473          0.0649          0.2055

Fixed effects:
                                    mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)                     -4.1401 0.3571    -4.8511  -4.1398    -3.4323 -4.1391   0
factor(group)1                  -0.7166 0.0480    -0.8114  -0.7164    -0.6229 -0.7161   0
centeredLatitude                 0.0736 0.0256     0.0227   0.0736     0.1246  0.0736   0
factor(group)1:centeredLatitude -0.0334 0.0028    -0.0389  -0.0333    -0.0279 -0.0333   0

Random effects:
Name          Model
 trial    IID model

Model hyperparameters:
                      mean   sd     0.025quant 0.5quant 0.975quant mode
Precision for trial 0.6693 0.2933 0.2461     0.6220   1.3742     0.5249

Expected number of effective parameters(std dev): 14.77(0.0746)
Number of equivalent replicates : 1.76

Marginal Likelihood:  -147.66
```

The posterior mean log-odds ratio (comparing the odds of TB among vaccinated versus not) decreases by approximately 0.03 for each unit difference from the average latitude.

Model 6 (C):



BCG Example: recap

- With this example we illustrated a few ways in which we could combine the data from the different studies.
  - Random effects: model heterogeneity
    - (example: no trivial variation in the response rates across studies!)
- Which model?
  - model choice guided by scientific questions
  - model choice guided by statistical criteria

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Markov Chain Monte Carlo (MCMC) Methods

(Implementation via JAGS)

Markov Chains

- Definition:
  - A **Markov Chain** is a sequence of random variables X_1, X_2, X_3, ... with the Markovian property, namely that, given the present state, the future and past states are independent. Formally,

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

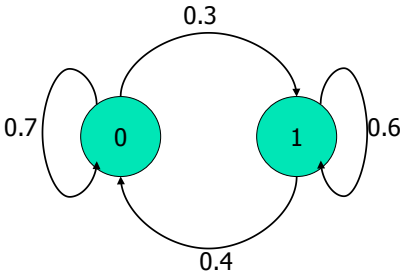
- Definition:
  - A Markov Chain is homogeneous if

$$P(X_{n+1} = y \mid X_n = x) = P(X_n = y \mid X_{n-1} = x) = P(x, y)$$

Markov Chains

- Example:
  - State Space: S={0,1}
  - Transition Matrix: (conditional probs. in rows)

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$



How does it behave?

Markov Chains

- Transition matrix in n steps?

$$P^n = S\Lambda^n S^{-1}$$

  - In our example, the eigenvalues of **P** are 1 and 0.3 with corresponding eigenvectors (1,1)' and (0.3,-0.4)'.
  - Thus:

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 0.3 \end{bmatrix}, S = \begin{bmatrix} 1 & 0.3 \\ 1 & -0.4 \end{bmatrix}, S^{-1} = \begin{bmatrix} 4/7 & 3/7 \\ 10/7 & -10/7 \end{bmatrix}$$

$$P^n = \begin{bmatrix} 4/7 + (0.3^{n+1})10/7 & 3/7 + (0.3^{n+1})10/7 \\ 4/7 - (0.3^n)4/7 & 3/7 + (0.3^n)4/7 \end{bmatrix}$$

Markov Chains

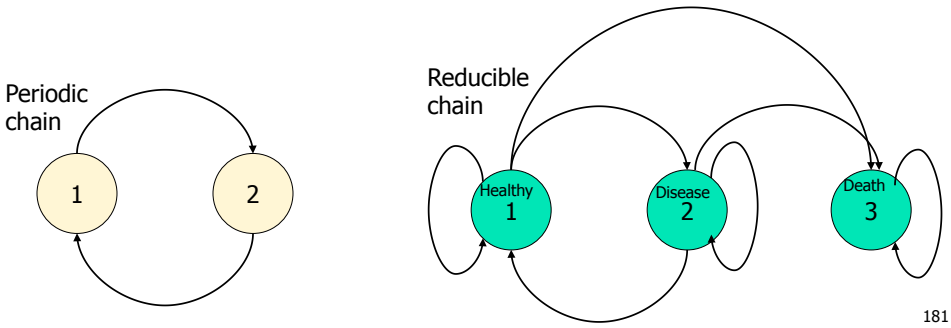
- Limiting distribution:

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 4/7 & 3/7 \\ 4/7 & 3/7 \end{bmatrix}$$

- Note that:
  - Largest eigenvalue is 1 (this gives the stationary distribution)
  - Rate of convergence is given by the second eigenvalue
  - Convergence describe "state" after many iterations
  - Stationary distribution does not depend on initial state
  - "Subliminal" message:
    - If we want to generate an observation from π, we can start anywhere and generate values from the transition probability matrix. After a length of time (burn-in), we can pick X_m whose distribution is π!

Markov Chains

- Conditions for convergence:
  - Aperiodic
    - Avoids the chain from oscillating between different sets in a regular movement
  - Irreducible
    - Starting from any point, the MC can reach any set with positive probability



Markov Chains and MCMC

- Q: How do we construct a Markov Chain whose stationary distribution is our target (posterior) distribution?
- A: Markov Chain Monte Carlo (MCMC)

Luckily, for most models, you can use existing software. Bugs/Winbugs/Jags are very popular. However, some models are more complex and you would need to implement your own MCMC (beyond the scope of this module)…

MCMC methods

- Implementing your own MCMC can be challenging!
- A large variety of models can be implemented in Bugs/ Winbugs/Jags
  - "Black-Box"
    - You will not need to derive full conditionals
    - You will not need to decide on MCMC samplers
  - Input:
    - Likelihood
    - Priors
    - [Define any quantity of interest (e.g. Odds Ratio, etc)]
  - Output
    - Posterior samples

Jags (Just Another Gibbs Sampler)

- Website:

  http://mcmc-jags.sourceforge.net

For MAC: http://sourceforge.net/projects/mcmc-jags/files/JAGS/3.x/Mac OS X/

- Very similar to WinBUGS (with a few differences)
- Goals/features:
  - Cross-platform engine for the BUGS language
  - Extensible, allowing users to write their own functions, distributions and samplers.
  - Platform for experimentation with ideas in Bayesian modelling
- Packages:
  - rjags: Allows you to run Jags from within R
  - coda: Allows you to perform convergence diagnosis

Bayesian estimation, inference and prediction using JAGS

Using Jags

Name	Usage	Density	Lower	Upper
Beta	`dbeta(a,b)` $a>0,\ b>0$	$\frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}$	0	1
Chi-square	`dchisqr(k)` $k>0$	$\frac{x^{\frac{k}{2}-1}\exp(-x/2)}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}$	0	
Double exponential	`ddexp(mu,tau)` $\tau>0$	$\tau\exp(-\tau\lvert x-\mu\rvert)/2$		
Exponential	`dexp(lambda)` $\lambda>0$	$\lambda\exp(-\lambda x)$	0	
F	`df(n,m)` $n>0,\ m>0$	$\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\left(\frac{n}{m}\right)^{\frac{n}{2}}x^{\frac{n}{2}-1}\left\{1+\frac{nx}{m}\right\}^{-\frac{(n+m)}{2}}$	0	
Gamma	`dgamma(r, lambda)` $\lambda>0,\ r>0$	$\frac{\lambda^r x^{r-1}\exp(-\lambda x)}{\Gamma(r)}$	0	
Generalized gamma	`dgen.gamma(r,lambda,b)` $\lambda>0,\ b>0,\ r>0$	$\frac{b\lambda^{br}x^{br-1}\exp\{-(\lambda x)^b\}}{\Gamma(r)}$	0	
Logistic	`dlogis(mu, tau)` $\tau>0$	$\frac{\tau\exp\{(x-\mu)\tau\}}{[1+\exp\{(x-\mu)\tau\}]^2}$		
Log-normal	`dlnorm(mu,tau)` $\tau>0$	$\left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}}x^{-1}\exp\left\{-\tau(\log(x)-\mu)^2/2\right\}$	0	
Noncentral Chi-squre	`dnchisqr(k, delta)` $k>0,\delta\geq 0$	$\sum_{r=0}^{\infty}\frac{\exp(-\frac{\delta}{2})(\frac{\delta}{2})^r}{r!}\frac{x^{(k/2+r-1)}\exp(-\frac{x}{2})}{2^{(k/2+r)}\Gamma(\frac{k}{2}+r)}$	0	
Normal	`dnorm(mu,tau)` $\tau>0$	$\left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}}\exp\{-\tau(x-\mu)^2/2\}$		
Pareto	`dpar(alpha, c)` $\alpha>0,\ c>0$	$\alpha c^{\alpha}x^{-(\alpha+1)}$	c	
Student t	`dt(mu,tau,k)` $\tau>0,\ k>0$	$\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\left(\frac{\tau}{k\pi}\right)^{\frac{1}{2}}\left\{1+\frac{\tau(x-\mu)^2}{k}\right\}^{-\frac{(k+1)}{2}}$		
Uniform	`dunif(a,b)` $a<b$	$\frac{1}{b-a}$	a	b
Weibull	`dweib(v, lambda)` $v>0,\ \lambda>0$	$v\lambda x^{v-1}\exp(-\lambda x^v)$	0	

Table 6.1: Univariate real-valued distributions in the bugs module

Using Jags

Name	Usage	Density	Lower	Upper
Beta	`dbetabin(a, b, n)`	$\binom{a+x-1}{x}\binom{b+n-x-1}{n-x}\binom{a+b+n-1}{n}^{-1}$	0	n
binomial	$a>0, b>0, n \in \mathbb{N}^*$			
Bernoulli	`dbern(p)`	$p^x(1-p)^{1-x}$	0	1
	$0<p<1$			
Binomial	`dbin(p,n)`	$\binom{n}{x}p^x(1-p)^{n-x}$	0	n
	$0<p<1, n \in \mathbb{N}^*$			
Categorical	`dcat(pi)`	$\frac{\pi_x}{\sum_i \pi_i}$	1	N
	$\pi \in (\mathbb{R}^+)^N$			
Noncentral	`dhyper(n1,n2,m1,psi)`	$\frac{\binom{n_1}{x}\binom{n_2}{m_1-x}\psi^x}{\sum_i \binom{n_1}{i}\binom{n_2}{m_1-i}\psi^i}$	$\max(0,n_+-m_1)$	$\min(n_1,m_1)$
hypergeometric	$0 \le n_i, 0 < m_1 \le n_+$			
Negative	`dnegbin(p, r)`	$\binom{x+r-1}{x}p^r(1-p)^x$	0	
binomial	$0<p<1, r \ge 0$			
Poisson	`dpois(lambda)`	$\frac{\exp(-\lambda)\lambda^x}{x!}$	0	
	$\lambda > 0$			

Table 6.2: Discrete univariate distributions in the bugs module

Name	Usage	Density
Dirichlet	`p ~ ddirch(alpha)`	$\Gamma(\sum_i \alpha_i)\prod_j \frac{p_j^{\alpha_j-1}}{\Gamma(\alpha_j)}$
	$\alpha_j \ge 0$	
Multivariate	`x ~ dmnorm(mu,Omega)`	$\left(\frac{\lvert\Omega\rvert}{2\pi}\right)^{\frac{1}{2}} exp\{-(x-\mu)^T\Omega(x-\mu)/2\}$
normal	Ω positive definite	
Wishart	`Omega ~ dwish(R,k)`	$\frac{\lvert\Omega\rvert^{(k-p-1)/2}\lvert R\rvert^{k/2}\exp\{-\mathrm{Tr}(R\Omega/2)\}}{2^{pk/2}\Gamma_p(k/2)}$
	R $p \times p$ pos. def., $k \ge p$	
Multivariate	`x ~ dmt(mu,Omega,k)`	$\frac{\Gamma\{(k+p)/2\}}{\Gamma(k/2)(n\pi)^{p/2}}\lvert\Omega\rvert^{1/2}\{1+\frac{1}{k}(x-\mu)^T\Omega(x-\mu)\}^{-\frac{(k+p)}{2}}$
Student t	Ω pos. def.	
Multinomial	`x ~ dmulti(pi, n)`	$n!\prod_j \frac{\pi_j^{x_j}}{x_j!}$
	$\sum_j x_j = n$	

Table 6.3: Multivariate distributions in the bugs module

Link function	Description	Range	Inverse
`cloglog(y) <- x`	Complementary log log	$0<y<1$	`y <- icloglog(x)`
`log(y) <- x`	Log	$0<y$	`y <- exp(x)`
`logit(y) <- x`	Logit	$0<y<1$	`y <- ilogit(x)`
`probit(y) <- x`	Probit	$0<y<1$	`y <- phi(x)`

Table 5.4: Link functions in the bugs module

Function	Description	Restrictions
`inprod(x1,x2)`	Inner product	Dimensions of $x1$, $x2$ conform
`interp.lin(e,v1,v2)`	Linear Interpolation	e scalar, $v1, v2$ conforming vectors
`logdet(m)`	Log determinant	m is a symmetric positive definite mat
`max(x1,x2,...)`	Maximum element among all arguments	
`mean(x)`	Mean of elements of x	
`min(x1,x2,...)`	Minimum element among all arguments	
`prod(x)`	Product of elements of x	
`sum(x)`	Sum of elements of x	
`sd(x)`	Standard deviation of elements of x	

Table 5.5: Scalar-valued functions with general arguments in the bugs module

Usage	Description	Restrictions
`inverse(a)`	Matrix inverse	a is a symmetric positive definite matrix
`rank(v)`	Ranks of elements of v	v is a vector
`order(v)`	Ordering permutation of v	v is a vector
`sort(v)`	Elements of v in order	v is a vector
`t(a)`	Transpose	a is a matrix
`a %*% b`	Matrix multiplication	a, b conforming vector or matrices

Table 5.6: Vector- or matrix-valued functions in the bugs module

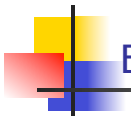
Example 1: using jags

```
model{
  ## define likelihood of observations
  for (i in 1:n){
     y[i] ~ dnorm(mu, tausq)
  }
  ## define priors
  mu ~ dnorm(0.0, 0.0001)
  tausq <- 1/sigmasq
  sigmasq ~ dunif(0,100)
}
```

Code saved in a text file
(in this case, example1.jag)

Example 1: using jags

```
## simulate data
data.sim <- function(N=1000, mu=0, sigma2=.5){
  y = rnorm(N, mu, sqrt(sigma2))
  return(y)
}

## true values for simulation
n       <- 100
mu      <-0
sigmasq<- 5

## simulated data
set.seed(1)
y <- data.sim(N=n, mu=mu, sigma2=sigmasq)

## load libraries
library(coda)
library(rjags)

## now prepare data for Jags
data <- list(y=y, n=n)

## initial values
inits <- list(mu=0, sigmasq=1)
```



Example 1: using jags

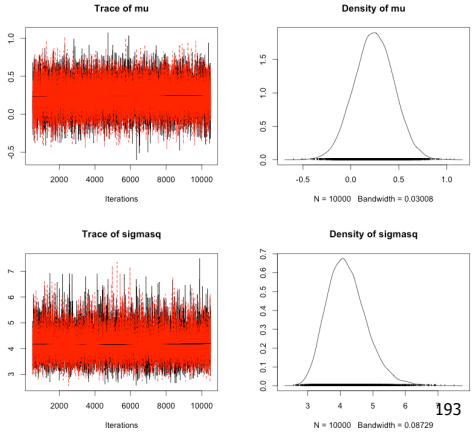
```
## define jags model within R
jags.m <- jags.model(file="example1.jag", data=data, inits=inits,
           n.chains=2, n.adapt=500)

## specify parameters to be monitored
params <- c("mu","sigmasq")

## run jags and save posterior samples
samps <- coda.samples(jags.m, params, n.iter=10000)

## summarize posterior samples
summary(samps)
summary(window(samps, start=1000))
plot(samps)
```


Example 1: using jags

```
> jags.m <- jags.model(file="example1.jag", data=data, inits=inits, n.chains=2, n.adapt=500)
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
   Graph Size: 108
Initializing model
  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100%
> ## specify parameters to be monitored
> params <- c("mu","sigmasq")
> ## run jags and save posterior samples
> samps <- coda.samples(jags.m, params, n.iter=10000)
  |**************************************************| 100%
> ## summarize posterior samples
> summary(samps)

Iterations = 501:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
          Mean     SD Naive SE Time-series SE
mu      0.2417 0.2057 0.001454       0.001454
sigmasq 4.2044 0.6123 0.004329       0.005847

2. Quantiles for each variable:
           2.5%    25%   50%    75%  97.5%
mu      -0.1595 0.1037 0.243 0.3812 0.6408
sigmasq  3.1678 3.7724 4.150 4.5722 5.5553
```

Example 1: using jags

```
> summary(window(samps, start=1000))

Iterations = 1000:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 9501

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

         Mean     SD Naive SE Time-series SE
mu      0.242 0.2057 0.001492       0.001492
sigmasq 4.207 0.6138 0.004453       0.006198

2. Quantiles for each variable:

           2.5%    25%    50%    75%  97.5%
mu      -0.1594 0.1034 0.2434 0.3815 0.6401
sigmasq  3.1672 3.7730 4.1526 4.5749 5.5636
```



Convergence Diagnostics Methods

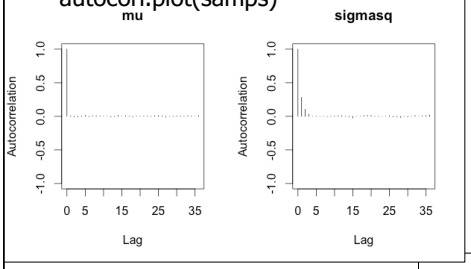
- Brooks, Gelman & Rubin
  - Two or more parallel chains (different starting values)
  - Comparison of within and between chain variance for each variable using the second half of chains
  - *"Rule-of-thumb": Samples are considered to arise from the stationary distribution if estimates are approximately equal to 1 (0.975 quantile is less than or equal to 1.2)*
- Geweke
  - Individual chain
  - Chain divided in two "windows" – comparison of the mean of sampled values in the first window to the mean in the second window
  - *"Rule-of-thumb": Lack of convergence if p-values < 0.05*

Convergence Diagnostics Methods

- Heidelberger and Welch
  - Individual chains
  - Based on Brownian bridge theory and uses Cramer-von-Mises statistic
  - Repeatedly discards 10% of iterations until the chain passes the test, or more than 50% of the iterations have been discarded
  - *"Rule-of-Thumb": Failure of the chain to pass the test indicates that a longer run is needed*
- Raftery and Lewis
  - Individual chains
  - *"Rule-of-Thumb": Dependence factors greater than 5 indicate lack of convergence*

Example 1: using jags

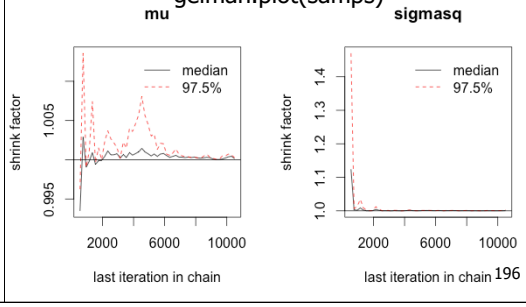
autocorr.plot(samps)

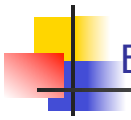


gelman.diag(samps)

```
Potential scale reduction factors:
        Point est. Upper C.I.
mu                1          1
sigmasq           1          1

Multivariate psrf
1
```

gelman.plot(samps)





Example 1: using jags

geweke.diag(samps)

```
[[1]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

     mu sigmasq
-0.4963 -0.6335


[[2]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

     mu sigmasq
-0.2554  0.2781
```

raftery.diag(samps)

```
[[1]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

         Burn-in  Total Lower bound  Dependence
         (M)      (N)    (Nmin)        factor (I)
 mu      2        3865  3746          1.03
 sigmasq 4        5299  3746          1.41


[[2]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

         Burn-in  Total Lower bound  Dependence
         (M)      (N)    (Nmin)        factor (I)
 mu      2        3771  3746          1.01
 sigmasq 4        5210  3746          1.39
```


Example 1: using jags

heidel.diag(samps)

```
[[1]]

        Stationarity start     p-value
        test         iteration
mu      passed       1         0.503
sigmasq passed       1         0.533

        Halfwidth Mean  Halfwidth
        test
mu      passed    0.242 0.0040
sigmasq passed    4.210 0.0158

[[2]]

        Stationarity start     p-value
        test         iteration
mu      passed       1         0.563
sigmasq passed       1         0.259

        Halfwidth Mean  Halfwidth
        test
mu      passed    0.241 0.00406
sigmasq passed    4.199 0.01658
```

Example 1: using jags

Posterior predictive distribution

```
## adding observation at last position for prediction (value is missing with NA)
y <- c(y, NA)
n <- length(y)
data <- list(y=y, n=n)
inits <- list(mu=0, sigmasq=1)
jags.m <- jags.model(file="example1.jag", data=data, inits=inits, n.chains=2, n.adapt=500)
params <- c("mu","sigmasq","y")
samps <- coda.samples(jags.m, params, n.iter=2000)
summary(samps)
```


Example 1: using jags

Posterior predictive distribution

```
> summary(samps)

Iterations = 501:2500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 2000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean     SD Naive SE Time-series SE
mu       0.243185 0.2036 0.003219       0.003219
sigmasq  4.199118 0.6070 0.009597       0.012994
y[1]    -1.400793 0.0000 0.000000       0.000000
y[2]     0.410639 0.0000 0.000000       0.000000
y[3]    -1.868522 0.0000 0.000000       0.000000
…
y[100]  -1.058556 0.0000 0.000000       0.000000
y[101]   0.297378 2.0684 0.032704       0.033294

2. Quantiles for each variable:

             2.5%       25%       50%       75%     97.5%
mu      -0.153727  0.111128  0.242856  0.374422  0.641681
sigmasq  3.183236  3.764582  4.134646  4.580882  5.517943
y[1]    -1.400793 -1.400793 -1.400793 -1.400793 -1.400793
…
y[100]  -1.058556 -1.058556 -1.058556 -1.058556 -1.058556
y[101]  -3.825483 -1.086550  0.326903  1.669352  4.398195
```

Revisiting the FEV Data Set with Jags

```
## Revisiting the analysis of the FEV dataset using Jags
## read FEV data set
data = read.table("data/fev.txt", col.names=c("seqnbr", "subjid", "age", "fev", "height", "sex", "smoke"))

## now prepare data for Jags
datajag <- list(n=length(data$fev), y=log(data$fev), smoke=1*(data$smoke==2), age=data$age-mean(data$age))
inits <- list(beta=rep(0,3),sigmasq=1)

## define jags model within R
model <- jags.model(file="fev.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
params <- c("beta", "sigmasq", "ratiogm")
fev.post <- coda.samples(model, params, n.iter=10000)

## summarize posterior samples
summary(fev.post)
plot(fev.post)

## convergence diagnosis
autocorr.plot(fev.post)
gelman.plot(fev.post)
heidel.diag(fev.post)
```

```
model{
  ## define likelihood of observations
  for (i in 1:n){
    y[i] ~ dnorm(mu[i], tausq)
    mu[i] <- beta[1] + beta[2]*smoke[i] + beta[3]*age[i]
  }
  ## define priors
  for (i in 1:3){
    beta[i] ~ dnorm(0, 0.0001)
  }
  tausq <- 1/sigmasq
  sigmasq ~ dunif(0,10)
  ## deriving quantities of interest (ratios of geometric means)
  for (i in 1:2){
    ratiogm[i] <- exp(beta[i+1])
  }
}
```


Revisiting the FEV Data Set with Jags

```
> summary(fev.post)

Iterations = 501:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean       SD  Naive SE Time-series SE
beta[1]    0.83441 0.028043 1.983e-04      9.359e-04
beta[2]    0.08997 0.029832 2.109e-04      1.004e-03
beta[3]    0.09078 0.003063 2.166e-05      4.416e-05
ratiogm[1] 1.09463 0.032622 2.307e-04      1.096e-03
ratiogm[2] 1.09503 0.003354 2.371e-05      4.836e-05
sigmasq    0.04469 0.002473 1.749e-05      2.181e-05

2. Quantiles for each variable:

              2.5%     25%     50%     75%   97.5%
beta[1]    0.78038 0.81522 0.83430 0.85315 0.89011
beta[2]    0.03107 0.07004 0.09003 0.11047 0.14729
beta[3]    0.08472 0.08872 0.09077 0.09285 0.09678
ratiogm[1] 1.03156 1.07255 1.09420 1.11680 1.15869
ratiogm[2] 1.08841 1.09277 1.09502 1.09730 1.10162
sigmasq    0.04010 0.04297 0.04462 0.04632 0.04972
```

Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons

Hwanhee Hong, MS, Bradley P. Carlin, PhD, Tatyana A. Shamliyan, MD, MS, Jean F. Wyman, PhD, Rema Ramakrishnan, MPH, François Sainfort, PhD, Robert L. Kane, MD

Objectives. *Bayesian statistical methods are increasingly popular as a tool for meta-analysis of clinical trial data involving both direct and indirect treatment comparisons. However, appropriate selection of prior distributions for unknown model parameters and checking of consistency assumptions required for modeling remain particularly challenging. We compared Bayesian and traditional frequentist statistical methods for mixed treatment comparisons with multiple binary outcomes.* ***Data***. *We searched major electronic bibliographic databases, Food and Drug Administration reviews, trial registries, and research grant databases up to December 2011 to find randomized studies published in English that examined drugs for female urgency urinary incontinence (UI) on continence, improvement in UI, and treatment discontinuation due to harm.* ***Methods***. *We describe and fit fixed and random effects models in both Bayesian and frequentist statistical frameworks. In a hierarchical model of 8 treatments, we separately analyze 1 safety and 2 efficacy outcomes. We produce Bayesian and frequentist treatment ranks and odds ratios across all drug v placebo comparisons, as well as Bayesian probabilities that each drug is best overall through a weighted scoring rule that trades off efficacy and safety.* ***Results***. *In our study, Bayesian and frequentist random effects models generally suggest the same drugs as most attractive, although neither suggests any significant differences between drugs. However, the Bayesian methods more consistently identify one drug (propiverine) as best overall, produce interval estimates that are generally better at capturing all sources of uncertainty in the data, and also permit attractive "rankograms" that visually capture the probability that each drug assumes each possible rank.* ***Conclusions***. *Bayesian methods are more flexible and their results more clinically interpretable, but they require more careful development and specialized software.* ***Key words:*** *nephrology; Bayesian meta-analysis; comparative effectiveness; systematic reviews; hierarchical models.* ***(Med Decis Making 2013;33: 702–714)***

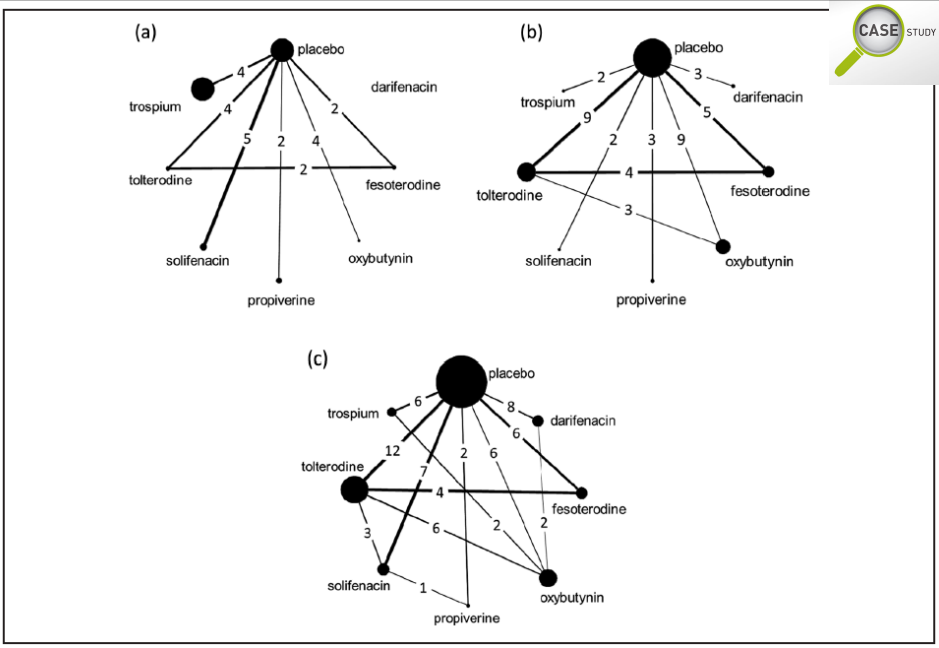


Figure 1 Network graphs of urinary incontinence (UI) data for each outcome: (a) continence, (b) UI improvement, and (c) discontinuation due to adverse events (AEs). The size of each node represents the number of studies investigating the drug, and the thickness of each edge implies the total number of samples for the relation. The number on the line is the number of studies for the relation.

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Suppose there are several trials
  - Comparing treatment A to B (AB trials)
    - Trials AB provide "direct evidence" of the effect of treatment B relative to A.
  - Comparing treatment A to C (AC trials)
    - Trials AC provide "direct evidence" of the effect of treatment C relative to A.
  - Comparing treatment B to C (BC trials)
    - Trials BC provide "direct evidence" of the effect of treatment C relative to B.

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Suppose there are several trials
  - What if: NO LONGER TRIALS AB!!!
  - Comparing treatment A to C (AC trials)
    - Trials AC provide "direct evidence" of the effect of treatment C relative to A.
  - Comparing treatment B to C (BC trials)
    - Trials BC provide "direct evidence" of the effect of treatment C relative to B.

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Best evidence on the effect of treatment B relative to A is provided by head-to-head trials.
- In the absence (or even sparsity) of such trials, there can be "indirect" evidence of the effect of B relative to A:

$$d_{AB}^{indirect} \overset{???}{=} d_{BC}^{direct} - d_{AC}^{direct}$$

- The mixing of direct and indirect evidence is called "mixed treatment comparison" (MTC)

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- More generically:
  - With K treatments, there are a total of K(K-1)/2 possible pairwise comparisons
    - E.g. K=6 means 15 potential comparisons of interest
  - Direct evidence for a subset of pairwise comparisons
  - Extending (pairwise) meta-analysis for MTD
    - Fixed effects model
    - Random effects model

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
> data= read.csv("mtc.csv")
> head(data, 20)
   Study Treatment Response    N Baseline
1      1         1        9  140        1
2      1         3       23  140        1
3      1         4       10  138        1
4      2         2       11   78        2
5      2         3       12   85        2
6      2         4       29  170        2
7      3         1       75  731        1
8      3         3      363  714        1
9      4         1        2  106        1
10     4         3        9  205        1
11     5         1       58  549        1
12     5         3      237 1561        1
13     6         1        0   33        1
14     6         3        9   48        1
15     7         1        3  100        1
16     7         3       31   98        1
17     8         1        1   31        1
18     8         3       26   95        1
19     9         1        6   39        1
20     9         3       17   77        1
…
```

Data from a Smoking Cessation Study

Randomized trials: 24 RCTs

Interventions:
- A: No Contact
- B: Self-Help
- C: Individual Counseling
- D: Group Counseling

Response:
Number of patients ceasing smoking

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Four Treatments:
  - A (reference)
  - B
  - C
  - D
- Direct evidence for:

$$d_{AB}, d_{AC}, d_{AD}$$

(basic parameters)

- Total number of contrasts: 6

- Indirect evidence for:

$$d_{BC} = d_{AC} - d_{AB}$$
$$d_{BD} = d_{AD} - d_{AB}$$
$$d_{CD} = d_{AD} - d_{AC}$$

- Consistency:
  - "Rationale":
    If (b-a)=2, (c-a)=3, then (c-b) must be 1

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Fixed Effects

$$r_{jk} \sim Binomial(p_{jk}, N_{jk})$$
$$\text{logit}(p_{jk}) = \mu_j + d_{XY} I_{(k=Y)}$$
$$d_{BC} = d_{AC} - d_{AB}$$
$$d_{BD} = d_{AD} - d_{AB}$$
$$d_{CD} = d_{AD} - d_{AC}$$
$$\mu_j, d_{AB}, d_{AC}, d_{AD} \sim N(0, 100^2)$$

Treatment effect in the baseline group for study j

Effect of treatment Y relative to X in trial j
(Y and X in generic notation)

- Random Effects

$$r_{jk} \sim Binomial(p_{jk}, N_{jk})$$
$$\text{logit}(p_{jk}) = \mu_j + \delta_{jXY} I_{(k=Y)}$$
$$\delta_{jXY} \sim N(d_{XY}, \sigma^2)$$
$$d_{BC} = d_{AC} - d_{AB}$$
$$d_{BD} = d_{AD} - d_{AB}$$
$$d_{CD} = d_{AD} - d_{AC}$$
$$\mu_j, d_{AB}, d_{AC}, d_{AD} \sim N(0, 100^2)$$
$$\sigma \sim U(0, 2)$$

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

Preparing for Coding in Jags:

Treatment Contrast	Treatment[i]	Baseline[i]	d[Treatment[i]]-d[Baseline[i]]	
1,2	2	1	d[2]-d[1]=d[2]	d_{AB}
1,3	3	1	d[3]-d[1]=d[3]	d_{AC}
1,4	4	1	d[4]-d[1]=d[4]	d_{AD}
2,3	3	2	d[3]-d[2]	$d_{BC} = d_{AC} - d_{AB}$
2,4	4	2	d[4]-d[2]	$d_{BD} = d_{AD} - d_{AB}$
3,4	4	3	d[4]-d[3]	$d_{CD} = d_{AD} - d_{AC}$

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

- Sometimes it is useful to have the absolute risk difference instead of odds ratios…
  - Can get this from (log-) odds ratios but need information about the "baseline" probability of the outcome:
    - What is the probability of smoking cessation in the "no treatment" group?
      - Can get this information from cohort studies, trials, etc
      - Assume, for example, that for "no treatment", the log-odds of smoking cessation has $N(-2.6, 0.38^2)$ distribution
      - Absolute effects for other treatments are:
        - $\text{Logit}(T_k) = A + d_k$

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
## Fixed Effects model
model{
 #  loop over 50 observations
 for(i in 1:50) {
   # likelihood
   Response[i] ~ dbin(p[i],N[i])
                 logit(p[i]) <- mu[Study[i]]+ delta[i]*(1-equals(Treatment[i], Baseline[i]))
   delta[i] <- d[Treatment[i]] - d[Baseline[i]]
 }
 # vague priors for intercepts (effect for baseline comparison group)
 for (j in 1:24) { mu[j] ~ dnorm(0,.0001)}
 # set effect of Treatment 1 as 0 (effects of other Treatments is relative to this Treatment 1)
 d[1] <- 0
 # flat priors for 3 basic treatment effect parameters
 for (k in 2:4) {
    d[k] ~ dnorm(0,.001)
 }

 # Absolute treatment effects
 # prior precision for Treatment 1, sd=.38
 precA <- pow(.38,-2)
 # external info on A.
 A ~ dnorm(-2.6,precA)

 for (i in 1:4){
  logit(T[i]) <- A + d[i]
 }

 #Rank the treatment effects (with 1=best) & record the best treatment
 rk <- 5- rank(d)
 best <- equals(rk,1)

 #All pairwise log odds ratios and odds ratios (some of these calculations are redundant, but needed to run)
 for (c in 1:4) {
   for (k in 1:4) {
     lor[c,k] <- d[k] - d[c]
   }
 }
}
```

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
# Random effects model
model{
  #  loop over 50 observations
  for(i in 1:50) {
    # likelihood
    Response[i] ~ dbin(p[i],N[i])
    logit(p[i]) <- mu[Study[i]]+ delta[i]*(1-equals(Treatment[i], Baseline[i]))

    # draw effect from random effects distribution
    delta[i] ~ dnorm(md[i],tau)

    # population mean effect
    md[i] <- d[Treatment[i]] - d[Baseline[i]]
  }
  # vague priors for intercepts (baseline group)
  for (j in 1:24) {
    mu[j] ~ dnorm(0,.0001)
  }
  # set effect of treatment 1 as 0 (all other treatment effects are relative to this one)
  d[1] <- 0
  # flat priors for 3 basic treatment parameters
  for (k in 2:4) {
    d[k] ~ dnorm(0,.001)
  }
  # Absolute treatment effects
  # prior precision for Treatment 1, sd=.38
  precA <- pow(.38,-2)
  # external info on A.
  A ~ dnorm(-2.6,precA)
  for (i in 1:4){
   logit(T[i]) <- A + d[i]
  }
  ## prior for variance component
  tau <- pow(sd,-2)
  sd ~ dunif(0,2)
  #Rank the treatment effects (with 1=best) & record the best treatment
  rk <- 5- rank(d)
  best <- equals(rk,1)
  #All pairwise log odds ratios and odds ratios (some of these calculations are redundant, but needed to run)
  for (c in 1:4) {
    for (k in 1:4) {
      lor[c,k] <- d[k] - d[c]
    }
    }
}
```


Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
## define jags model within R
datajag <- list(N=data$N, Study=data$Study, Response=data$Response, Treatment=data$Treatment, Baseline=data$Baseline)
inits <- list(A=1)
model1 <- jags.model(file="mtc-fe.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
params1 <- c("d", "lor", "rk", "best","T")
post1 <- coda.samples(model1, params1, n.iter=10000)
summary(post1)

## define jags model within R
datajag <- list(N=data$N, Study=data$Study, Response=data$Response, Treatment=data$Treatment, Baseline=data$Baseline)
inits <- list(A=1)
model2 <- jags.model(file="mtc-re.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
params2 <- c("d", "lor", "rk", "best","T",”sd”)
post2 <- coda.samples(model2, params2, n.iter=10000)
summary(post2)
```

Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
> summary(post1)

Iterations = 501:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean      SD  Naive SE Time-series SE
T[1]      0.07296 0.02620 0.0001853      0.0001853
T[2]      0.09022 0.03341 0.0002363      0.0002988
T[3]      0.14352 0.04735 0.0003348      0.0003539
T[4]      0.15490 0.05472 0.0003869      0.0005126
best[1]   0.00000 0.00000 0.0000000      0.0000000
best[2]   0.00000 0.00000 0.0000000      0.0000000
best[3]   0.31590 0.46488 0.0032872      0.0063026
best[4]   0.68410 0.46488 0.0032872      0.0063026
d[1]      0.00000 0.00000 0.0000000      0.0000000
d[2]      0.22728 0.12619 0.0008923      0.0024422
d[3]      0.76522 0.05784 0.0004090      0.0011966
d[4]      0.84744 0.17441 0.0012333      0.0030385
lor[1,1]  0.00000 0.00000 0.0000000      0.0000000
lor[2,1] -0.22728 0.12619 0.0008923      0.0024422
lor[3,1] -0.76522 0.05784 0.0004090      0.0011966
lor[4,1] -0.84744 0.17441 0.0012333      0.0030385
lor[1,2]  0.22728 0.12619 0.0008923      0.0024422
lor[2,2]  0.00000 0.00000 0.0000000      0.0000000
lor[3,2] -0.53795 0.13485 0.0009536      0.0025670
lor[4,2] -0.62016 0.19335 0.0013672      0.0034821
lor[1,3]  0.76522 0.05784 0.0004090      0.0011966
lor[2,3]  0.53795 0.13485 0.0009536      0.0025670
lor[3,3]  0.00000 0.00000 0.0000000      0.0000000
lor[4,3] -0.08222 0.17194 0.0012158      0.0027551
lor[1,4]  0.84744 0.17441 0.0012333      0.0030385
lor[2,4]  0.62016 0.19335 0.0013672      0.0034821
lor[3,4]  0.08222 0.17194 0.0012158      0.0027551
lor[4,4]  0.00000 0.00000 0.0000000      0.0000000
rk[1]     3.96280 0.18926 0.0013382      0.0025080
rk[2]     3.03660 0.19095 0.0013502      0.0025454
rk[3]     1.68410 0.46488 0.0032872      0.0063026
rk[4]     1.31650 0.46641 0.0032980      0.0063417
```

```
> summary(post2)

Iterations = 501:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean      SD  Naive SE Time-series SE
T[1]      0.07306 0.02630 0.0001860      0.0001860
T[2]      0.12326 0.05980 0.0004228      0.0006149
T[3]      0.15244 0.05826 0.0004120      0.0005674
T[4]      0.21058 0.09726 0.0006877      0.0012472
best[1]   0.00000 0.00000 0.0000000      0.0000000
best[2]   0.04945 0.21681 0.0015331      0.0018612
best[3]   0.18735 0.39020 0.0027591      0.0043136
best[4]   0.76320 0.42513 0.0030061      0.0048088
d[1]      0.00000 0.00000 0.0000000      0.0000000
d[2]      0.53029 0.39518 0.0027944      0.0051176
d[3]      0.81824 0.23657 0.0016728      0.0036416
d[4]      1.18190 0.46436 0.0032835      0.0073618
lor[1,1]  0.00000 0.00000 0.0000000      0.0000000
lor[2,1] -0.53029 0.39518 0.0027944      0.0051176
lor[3,1] -0.81824 0.23657 0.0016728      0.0036416
lor[4,1] -1.18190 0.46436 0.0032835      0.0073618
lor[1,2]  0.53029 0.39518 0.0027944      0.0051176
lor[2,2]  0.00000 0.00000 0.0000000      0.0000000
lor[3,2] -0.28795 0.40115 0.0028366      0.0048623
lor[4,2] -0.65161 0.48342 0.0034183      0.0065579
lor[1,3]  0.81824 0.23657 0.0016728      0.0036416
lor[2,3]  0.28795 0.40115 0.0028366      0.0048623
lor[3,3]  0.00000 0.00000 0.0000000      0.0000000
lor[4,3] -0.36366 0.45632 0.0032267      0.0060984
lor[1,4]  1.18190 0.46436 0.0032835      0.0073618
lor[2,4]  0.65161 0.48342 0.0034183      0.0065579
lor[3,4]  0.36366 0.45632 0.0032267      0.0060984
lor[4,4]  0.00000 0.00000 0.0000000      0.0000000
rk[1]     3.91070 0.29314 0.0020728      0.0027927
rk[2]     2.76745 0.66498 0.0047021      0.0071024
rk[3]     2.02115 0.62930 0.0044498      0.0065797
rk[4]     1.30070 0.58881 0.0041635      0.0066858
sd        0.83107 0.18728 0.0013243      0.0067195
```


Meta-Analysis:
Mixed and Indirect Treatment Comparisons

(model 1)

2. Quantiles for each variable:

```
             2.5%      25%      50%      75%    97.5%
T[1]      0.03338  0.05401  0.06931  0.08751  0.13360
T[2]      0.04018  0.06597  0.08526  0.10886  0.16894
T[3]      0.06876  0.10912  0.13785  0.17153  0.25160
T[4]      0.06965  0.11441  0.14775  0.18683  0.28209
best[1]   0.00000  0.00000  0.00000  0.00000  0.00000
best[2]   0.00000  0.00000  0.00000  0.00000  0.00000
best[3]   0.00000  0.00000  0.00000  1.00000  1.00000
best[4]   0.00000  0.00000  1.00000  1.00000  1.00000
d[1]      0.00000  0.00000  0.00000  0.00000  0.00000
d[2]     -0.02449  0.14317  0.22808  0.31171  0.47287
d[3]      0.65205  0.72584  0.76562  0.80482  0.87714
d[4]      0.50183  0.73074  0.84815  0.96487  1.18863
lor[1,1]  0.00000  0.00000  0.00000  0.00000  0.00000
lor[2,1] -0.47287 -0.31171 -0.22808 -0.14317  0.02449
lor[3,1] -0.87714 -0.80482 -0.76562 -0.72584 -0.65205
lor[4,1] -1.18863 -0.96487 -0.84815 -0.73074 -0.50183
lor[1,2] -0.02449  0.14317  0.22808  0.31171  0.47287
lor[2,2]  0.00000  0.00000  0.00000  0.00000  0.00000
lor[3,2] -0.80440 -0.62754 -0.53865 -0.44647 -0.27571
lor[4,2] -1.00369 -0.74952 -0.61940 -0.48892 -0.23973
lor[1,3]  0.65205  0.72584  0.76562  0.80482  0.87714
lor[2,3]  0.27571  0.44647  0.53865  0.62754  0.80440
lor[3,3]  0.00000  0.00000  0.00000  0.00000  0.00000
lor[4,3] -0.41659 -0.19782 -0.08315  0.03212  0.25543
lor[1,4]  0.50183  0.73074  0.84815  0.96487  1.18863
lor[2,4]  0.23973  0.48892  0.61940  0.74952  1.00369
lor[3,4] -0.25543 -0.03212  0.08315  0.19782  0.41659
lor[4,4]  0.00000  0.00000  0.00000  0.00000  0.00000
rk[1]     3.00000  4.00000  4.00000  4.00000  4.00000
rk[2]     3.00000  3.00000  3.00000  3.00000  4.00000
rk[3]     1.00000  1.00000  2.00000  2.00000  2.00000
rk[4]     1.00000  1.00000  1.00000  2.00000  2.00000
```

(model 2)

2. Quantiles for each variable:

```
             2.5%      25%      50%      75%    97.5%
T[1]      0.03400  0.05420  0.06877  0.08744  0.1357
T[2]      0.04156  0.08041  0.11183  0.15366  0.2690
T[3]      0.06525  0.11029  0.14350  0.18408  0.2910
T[4]      0.06966  0.13885  0.19332  0.26563  0.4403
best[1]   0.00000  0.00000  0.00000  0.00000  0.0000
best[2]   0.00000  0.00000  0.00000  0.00000  1.0000
best[3]   0.00000  0.00000  0.00000  0.00000  1.0000
best[4]   0.00000  1.00000  1.00000  1.00000  1.0000
d[1]      0.00000  0.00000  0.00000  0.00000  0.0000
d[2]     -0.24088  0.27085  0.52436  0.77945  1.3193
d[3]      0.36985  0.66280  0.80985  0.96789  1.3099
d[4]      0.28324  0.87465  1.17071  1.47888  2.1348
lor[1,1]  0.00000  0.00000  0.00000  0.00000  0.0000
lor[2,1] -1.31928 -0.77945 -0.52436 -0.27085  0.2409
lor[3,1] -1.30986 -0.96789 -0.80985 -0.66280 -0.3699
lor[4,1] -2.13481 -1.47888 -1.17071 -0.87465 -0.2832
lor[1,2] -0.24088  0.27085  0.52436  0.77945  1.3193
lor[2,2]  0.00000  0.00000  0.00000  0.00000  0.0000
lor[3,2] -1.08952 -0.54691 -0.28594 -0.02666  0.5045
lor[4,2] -1.61722 -0.96223 -0.64931 -0.33448  0.3008
lor[1,3]  0.36985  0.66280  0.80985  0.96789  1.3099
lor[2,3] -0.50450  0.02666  0.28594  0.54691  1.0895
lor[3,3]  0.00000  0.00000  0.00000  0.00000  0.0000
lor[4,3] -1.28817 -0.65763 -0.36012 -0.06286  0.5276
lor[1,4]  0.28324  0.87465  1.17071  1.47888  2.1348
lor[2,4] -0.30075  0.33448  0.64931  0.96223  1.6172
lor[3,4] -0.52756  0.06286  0.36012  0.65763  1.2882
lor[4,4]  0.00000  0.00000  0.00000  0.00000  0.0000
rk[1]     3.00000  4.00000  4.00000  4.00000  4.0000
rk[2]     1.00000  2.00000  3.00000  3.00000  4.0000
rk[3]     1.00000  2.00000  2.00000  2.00000  3.0000
rk[4]     1.00000  1.00000  1.00000  1.00000  3.0000
sd        0.54041  0.69643  0.80651  0.93827  1.2658
```

American Journal of Epidemiology
Copyright © 1995 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved
Vol. 141, No. 3
Printed in U.S.A.

CASE STUDY

Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard

TABLE 1. Results of serologic and stool testing for *Strongyloides* infection on 162 Cambodian refugees arriving in Montreal, Canada, between July 1982 and February 1983

		Stool examination +	Stool examination −	
Serology	+	38	87	125
	−	2	35	37
		40	122	162

		Prior information		Stool examination alone		Serology alone		Both tests combined	
		Median	95% CI*	Median	95% CI	Median	95% CI	Median	95% CI
	π	0.50	0.03–0.98	0.74	0.41–0.98	0.80	0.23–0.99	0.76	0.52–0.91
Stool examination	S_1	0.24	0.07–0.47	0.30	0.21–0.47			0.31	0.22–0.44
	C_1	0.95	0.89–0.99	0.95	0.88–0.99			0.96	0.91–0.99
	PPV_1	0.84	0.10–1.00	0.95	0.74–1.00			0.98	0.88–1.00
	NPV_1	0.56	0.03–0.98	0.33	0.02–0.73			0.30	0.11–0.63
Serology	S_2	0.81	0.63–0.92			0.83	0.73–0.92	0.89	0.80–0.95
	C_2	0.72	0.31–0.96			0.58	0.22–0.94	0.67	0.36–0.95
	PPV_2	0.76	0.07–1.00			0.91	0.18–1.00	0.90	0.62–1.00
	NPV_2	0.78	0.08–1.00			0.44	0.03–0.94	0.70	0.28–0.92

* CI, credible interval.

Reproducing analyses: Using only one diagnostic test

- Recall: In the absence of 'gold standard' we only observe totals

TABLE 2. Observed and latent data in the case of one diagnostic test in the absence of a gold standard, presented in a 2 × 2 table

		Truth +	Truth −	
Test	+	Y_1	$a-Y_1$	a
	−	Y_2	$b-Y_2$	b
		Y_1+Y_2	$N-(Y_1+Y_2)$	N

Y_1 and Y_2 are latent/unobserved data

Reproducing analyses: Using only one diagnostic test

- Probability model for positive test result?
  - a ~ Binomial(N, PPT)
    - Where N is the total sample size (i.e. a+b)
    - PPT is the probability of a positive test

$$PPT = P(T+) = P(T+ \mid D)P(D) + P(T+ \mid D^c)P(D^c)$$
$$= S\pi + (1-C)(1-\pi)$$

(recall: S is sensitivity and C is the specificity)

Reproducing analyses: Using only one diagnostic test

Jags Code

```
model{
  ## model
  a ~ dbin(PPT, N)

  ## definition of probability of positive test
  PPT <- S*pi + (1-C)*(1-pi)

  ## priors
  S ~ dbeta(aS,bS)     # prior for sensitivity
  C ~ dbeta (aC, bC)   # prior for specificity
  pi ~ dbeta(api, bpi) # prior for prevalence

  ## computing probability of disease given test results
  pY1 <- pi*S/PPT
  pY2 <- pi*(1-S)/(1-PPT)

  ## simulating Y1, Y2
  Y1 ~ dbin(pY1,a)
  Y2 ~ dbin(pY2, N-a)
}
```

Note: original paper derived full conditionals that allows one to implement full MCMC (Gibbs Sampler) – but that is out of the scope of this introductory course.

Reproducing analyses: Using only one diagnostic test

```
data = list(N=162, a=40, Y1=NA, Y2=NA, api=1, bpi=1, aS=4.4,
            bS=13.31, aC=71.25, bC=3.75)

## Initial values
inits = function(){list(pi=0.5, S=0.9, C=0.8, Y1=10, Y2=10)}

## Model specification
jags.m =jags.model(file=diagnostic.jag, data=data, n.chains=2,
           n.adapt=1000, inits=inits())

## Parameters to be monitored
params = c("pi", "S", "C", "Y1", "Y2")

## Sampling
samps <- coda.samples(jags.m, params,  n.iter=5000, thin=5)

## Summarize posterior samples and save output results
aux <- summary(samps)
par(mfrow=c(3,2))
plot(samps)
output <- cbind(aux[[1]][,c(1,2)], aux[[2]][,c(1,3,5)])
```

```
 Mean           SD       2.5%         50%       97.5%
C   0.9469572  0.02796301  0.8771444  0.9516872   0.9862617
S   0.3128120  0.06713805  0.2093188  0.3023151   0.4756297
Y1 37.4385000  3.10580759 29.0000000 38.0000000  40.0000000
Y2 82.8770000 24.44454211 35.9750000 85.0000000 120.0000000
pi  0.7401364  0.16335664  0.4075330  0.7581660   0.9855292
```

Posterior Estimates

Final Comments

- There is 'art' in Bayesian Analysis

- Achieving 'mastery' requires practice!