DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Bayesian Methods for Clinical Research: Computational Methods and Applications

Outline

- Introduction
- Bayesian regression models
- Correlated data
- MCMC
- Meta-analysis
- Mixed treatment comparisons
- Diagnostic testing with missing gold standard

Overview of the Bayesian approach

- 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 to analysis:
  - Prior distribution: $P(\theta)$
  - Likelihood Function: $L(\theta|y) \propto P(y|\theta)$

The likelihood function

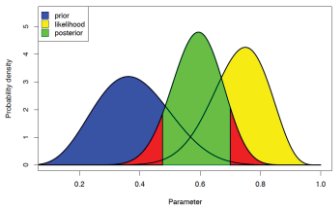
- A **likelihood function** is a function of the parameters of a probability model given the outcomes.
  - The *likelihood* of θ, given outcome y, is equal to the *probability* of that observed outcome given θ.
- For example, a Bernoulli random variable Y takes on two possible values: 0 or 1
  - Likelihood 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:
      - $L(\theta|y=0) = P(Y=0|\theta)=1-\theta$

Prior Distributions

- Quantifiable prior belief about the parameter of interest
- Choice of priors is based on judgments and a degree of subjectivity cannot be avoided
  - Prior is not unique!
  - Sensitivity analysis allows us to assess the impact of particular distributions on the conclusions
- Construct prior, $P(\theta)$, based on prior belief
  - Conjugate priors result in posterior in same family
  - Non-informative or flat prior

Posterior distribution

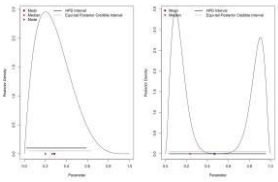
- Output from analysis:
  - Posterior distribution: $P(\theta|Y)=\frac{P(\theta)L(\theta|Y)}{\int P(\theta)L(\theta|Y)d\theta}$



Inference based on posterior distribution

- Inferences based on summaries of the <u>posterior distribution</u>
  - 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].

Primary packages we will use

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

```
> install.packages("INLA", repos=c(getOption("repos"), INLA="https://inla.r-inla-download.org/R/testing"), dep=TRUE)
> library(INLA)
```

- rjags (alternative choices R2jags, runjags)

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Bayesian Regression Models

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 INLA package
- Different approaches to estimation of GLMs
  - Approximate posterior inference (Bayesian CLT)
- Advantages:
  - Syntax very similar to 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

For more info: http://www.r-inla.org

Model	Name	Description
ndependent 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

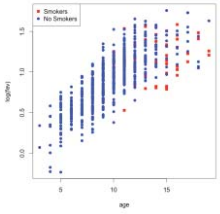
Negative Binomial	nbinomial	nbinomial.pdf
Poisson	poisson	poisson.pdf
Binomial	binomial	binomial.pdf
Cllinomial	cbinomial	cbinomial
Gaussian	gaussian	gaussian.pdf
Skew Normal	sn	sn.pdf
Student-t	T	Student-t.pdf
Gaussian model for stochastic volatility	stochvol	stochvolgaussian.pdf
Student-t model for 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 Log-gamma distribution	normal gaussian loggamma	gaussian.pdf prior-loggamma.pdf
Improper flat prior Truncated Normal distribution	flat logtnormal logtgaussian	prior-flat.pdf 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 Define your own prior	logitbeta expression:	logitbeta.pdf expression.pdf
Define your own prior	table:	table.pdf

Multiple Linear Regression in R

- FEV dataset: contains data on 654 children.
  https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/fev.txt
  - 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("https://raw.githubusercontent.com/rhubb/STS528317/master/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 :36871   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.   :90881   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: 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$summary.fix
                mean           sd  0.025quant    0.5quant   0.975quant        mode          kld
(Intercept) -0.15691451 0.079087993 -0.31222632 -0.15691674 -0.001736104 -0.15691451 4.714742e-14
smoke        0.08992701 0.031641008  0.02778011  0.08992612  0.152011540  0.08992701 4.721406e-14
age          0.09076807 0.003207235  0.08446975  0.09076798  0.097069977  0.09076807 4.216079e-14
> fit.prior1$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:

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

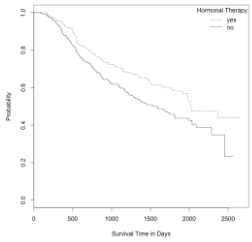
- Parametric vs Semi-parametric PH model?
  - What is the form of the baseline hazard ($h_0(.)$) function?

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).
- Scientific question of interest: does receipt of hormone therapy affect the length of time to breast cancer recurrence?

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
```

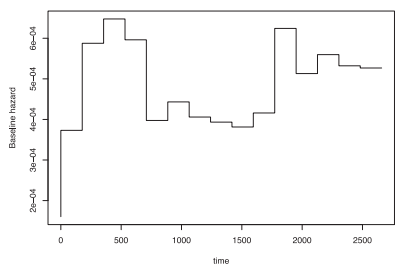
```
> ## Parametric survival (Weibull regression)
> library(eha)
> fit3 <- phreg(Surv(time, cens) ~ horTh, data=GBSG2, dist = "weibull")
> summary(fit3)
Call:
phreg(formula = Surv(time, cens) ~ horTh, data = GBSG2, dist = "weibull")

Covariate          W.mean      Coef Exp(Coef)  se(Coef)    Wald p
horTh
            no     0.634     0         1                (reference)
           yes     0.366    -0.393     0.675     0.125     0.002

log(scale)                   7.698                0.058     0.000
log(shape)                   0.251                0.050     0.000

Events                     299
Total time at risk         77140
Max. log. likelihood       -2632.1
LR test statistic          10.36
Degrees of freedom         1
Overall p-value            0.0012874
```

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) -7.7878 0.1483    -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  3962.23     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:
Non-parametric

```
> plot(fit$summary.random$baseline.hazard[,"ID"],
+      exp(fit$summary.fixed[1,1]+fit$summary.random$baseline.hazard[,"mean"]), type="S",
+      xlab = "time", ylab = "Baseline hazard")
```



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 2e-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
```

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY *of* WASHINGTON
School of Public Health

Analysis of Correlated Data

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

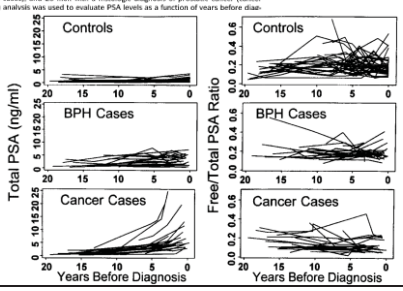
Example: Longitudinal data

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; srsubram@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.

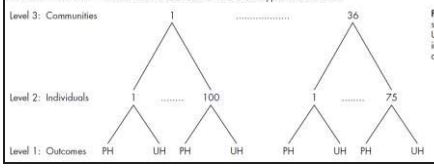


Figure 1 Multivariate multilevel structure of responses (PH, poor health; UH, unhappy) at level 1 nested within individuals at level 2 nested within communities at level 3

American Journal of Epidemiology
Vol. 145, No. 6
Printed in U.S.A.
Copyright © 1997 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

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

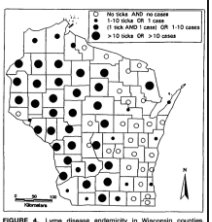


FIGURE 4. Lyme disease endemicity in Wisconsin counties, 1970–1990, as determined by county of exposure for human cases and known distribution of Ixodes scapularis.

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

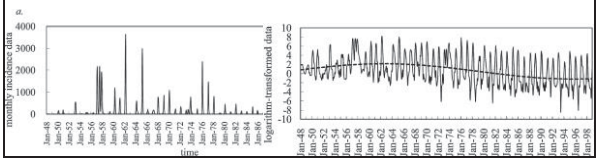
Background: Much effort has been expended on interpreting the mechanism of influenza epidemics, so as to better predict them. In addition to the obvious annual cycle of influenza epidemics, longer-term incidence patterns are present. These so-called interepidemic periods have long been a focus of epidemiology. 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

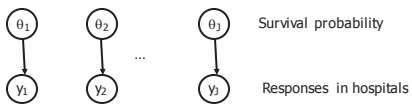


Modeling of Correlated Data:
Motivation

- 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:

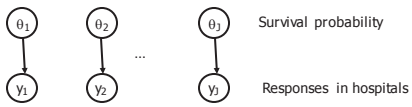
- Goal:
  - Study the effectiveness of cardiac treatments



Independent Data
(Separate analysis using data from each study)

Hierarchical Model Example:

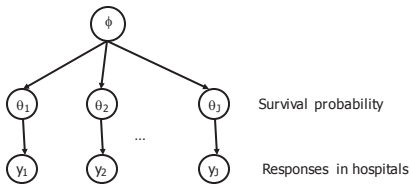
- Goal:
  - Study the effectiveness of cardiac treatments



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

Hierarchical Model Example:

- Goal:
  - Study the effectiveness of cardiac treatments



This implies, marginally, correlation between observations!

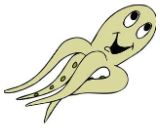
Hierarchical Model Example:

- 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:

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, <u>but</u> 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 $k \le n$,

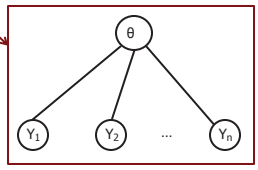
$$P[(k, n-k)] = \int_0^1 \theta^k (1-\theta)^{n-k}\, dF(\theta)$$

- 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

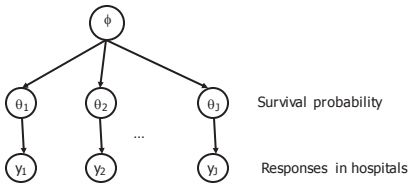
Representation:

$(Y_1,\ldots,Y_n)$ exchangeable



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 into 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?
    - …

Motivating example: HARVEST study

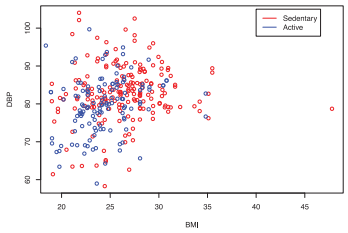
- We will use data from the HARVEST study to explore the use of INLA to estimate parameters in a hierarchical regression model
- Data are from 214 participants in a study of ambulatory blood pressure monitoring. Participants with suspected hypertension wore an ambulatory monitoring device for 24 hours to assess blood pressure on three occasions over the course of 5 months
  https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/harvest.csv

- The data set includes the following variables

  - ID: study id
  - Month: 1 = baseline, 3 = 3 month assessment, 5 = 5 month assessment
  - Smoke: number cigarettes smoked per day (0 = non-smoker, 1 = 1-5, 2 = 6-10, 3 = 11-20)
  - Sport: 0 = sedentary, 1 = light activity, 2 = non-competitive sports, 3 = competitive sports
  - SBP: systolic blood pressure (continuous)
  - DBP: diastolic blood pressure (continuous)
  - HR: heart rate (continuous)
  - Age: age in years at baseline (continuous)
  - BMI: body mass index at baseline (continuous)
  - Male: male = 1, female = 0

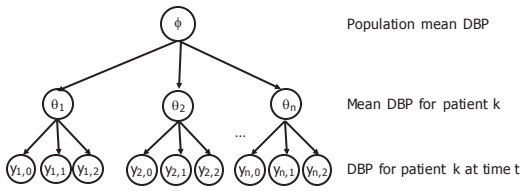
Motivating example: HARVEST study

- We would like to study the association between age, BMI, activity level, and diastolic blood pressure
- However, since multiple measurements are available for each subject it is necessary to account for correlation among DBP measures made for the same participant



Hierarchical Model for HAREST stude

- Let $y_{k,t}$= DBP for patient k at time t



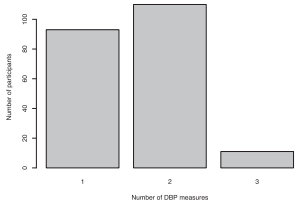
Population mean DBP

Mean DBP for patient k

DBP for patient k at time t

Motivating example: HARVEST study

```
> harvest <- read.csv("https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/harvest.csv",
header = T, na.strings = ".")
> summary(harvest[,c("SBP","DBP","Age","BMI","Male")])
      SBP             DBP              Age             BMI             Male
 Min.   :100.1   Min.   : 48.23   Min.   :16.00   Min.   :14.53   Min.   :0.000
 1st Qu.:123.9   1st Qu.: 77.92   1st Qu.:26.00   1st Qu.:22.85   1st Qu.:0.000
 Median :130.2   Median : 81.97   Median :33.00   Median :24.87   Median :1.000
 Mean   :131.1   Mean   : 81.49   Mean   :33.24   Mean   :25.36   Mean   :0.729
 3rd Qu.:137.8   3rd Qu.: 86.02   3rd Qu.:41.00   3rd Qu.:27.51   3rd Qu.:1.000
 Max.   :169.2   Max.   :104.05   Max.   :53.00   Max.   :47.80   Max.   :1.000
 NA's   :296     NA's   :296      NA's   :3       NA's   :48
```



Motivating example: HARVEST study

```
> mod <- inla(DBP ~ factor(Sport) + Age + BMI + f(ID, model = "iid"), family="gaussian", data = harvest)
> summary(mod)

Call:
c("inla(formula = DBP ~ factor(Sport) + Age + BMI + f(ID, model = \"iid\"), ",  "    family = \"gaussian\",
data = harvest)")

Time used:
 Pre-processing     Running inla Post-processing           Total
         0.6894           0.6137           0.0748          1.3780

Fixed effects:
                  mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)    79.0616 2.8079     73.5402  79.0625     84.5726 79.0644   0
factor(Sport)1 -1.9735 1.4128     -4.7531  -1.9725      0.7982 -1.9705   0
factor(Sport)2 -2.9389 1.5057     -5.8948  -2.9402      0.0208 -2.9426   0
factor(Sport)3 -4.5822 1.8989     -8.3137  -4.5825     -0.8530 -4.5829   0
Age             0.1617 0.0607      0.0424   0.1617      0.2809  0.1617   0
BMI            -0.0850 0.0700     -0.2226  -0.0850      0.0525 -0.0850   0

Random effects:
Name   Model
 ID    IID model

Model hyperparameters:
                                          mean     sd 0.025quant 0.5quant 0.975quant   mode
Precision for the Gaussian observations 0.0422 0.0023     0.0377    0.0422     0.0468 0.0421
Precision for ID                        0.0285 0.0029     0.0230    0.0285     0.0343 0.0284
```

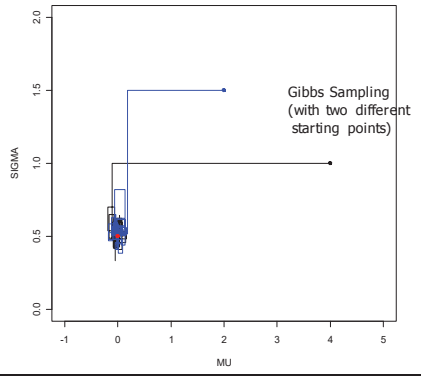
Precision of DBP across participants

Markov Chain Monte Carlo (MCMC)

(Implementation via JAGS)

MCMC algorithms

- Class of algorithms to sample from a target probability distribution
  - Construct a Markov chain for which the target distribution is its equilibrium distribution
  - Allow for sampling of the target distribution by observing the chain after a number of steps (burn-in)
- There are many MCMC algorithms
  - Gibbs Sampler
  - Metropolis Hastings
  - Hybrid MCMC
  - Adaptive MCMC
  - Slice Sampler
  - Reversible Jump MCMC (RJMCMC)
  - Particle filters
  - …
- Only discussing the simplest algorithm



Gibbs Sampling

- Derive full conditional distributions (distribution of each parameter conditional on all other parameters and data): $[\theta_i \mid \theta_{-i}]$
- Algorithm:
  1. Set starting values $(\theta_1^0, \ldots, \theta_k^0)$
  2. Iteration i
     a. Draw $\theta_1^{(i)} \sim [\theta_1 \mid \theta_2^{(i-1)}, \ldots, \theta_k^{(i-1)}]$
     b. Draw $\theta_2^{(i)} \sim [\theta_2 \mid \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_k^{(i-1)}]$
     c. …
     d. Draw $\theta_k^{(i)} \sim [\theta_k \mid \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{k-1}^{(i)}]$
  3. Repeat step 1. After t iterations obtain

Full conditionals

- Theorem: $(\theta_1^{(t)}, \ldots, \theta_k^{(t)}) \xrightarrow{D} [\theta_1, \ldots, \theta_k] \;\; as \;\; t \to \infty$

Gibbs Sampling

- Note:
  - You should be able to sample from the full conditionals.
    - If the full conditional is not in closed form, may need to use other algorithms (beyond the scope of today's lecture)…
  - In some problems, full conditionals are also complex…

Example 3:

Normal model with noninformative priors

Let $Y_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$ and $\pi(\mu, \sigma^2) \propto \sigma^{-2}$

We have:

$$\pi(\mu, \sigma^2 \mid y) \propto \sigma^{-2\left(\frac{n}{2}+1\right)} \exp\left\{-\frac{\sum(y_i - \mu)^2}{2\sigma^2}\right\}$$

Let $\tau = \sigma^{-2}$. One can show

$$\pi(\mu \mid \sigma^2, y) = N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

$$\pi(\tau \mid \mu, y) = Gamma\left(\frac{n}{2} + 2, \frac{\sum(y_i - \mu)^2}{2}\right)$$

Example 3:

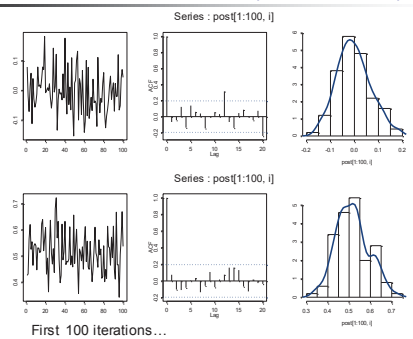
Normal model with noninformative priors

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

## Full conditionals implementation #########
fc.mu <- function(y, sigma2=0.1){
  mu = rnorm(1, mean(y), sqrt(sigma2/length(y)))
  return(mu)
}

fc.sigma2 <- function(y, mu=0){
  tau = rgamma(1, (length(y)/2)+2, rate=sum((y-mu)^2)/2)
  sigma2= 1/tau
  return(sigma2)
}
```

```
## Main program (GIBBS) #########
y = data.sim(N=100, mu=0, sigma2=0.5)

# length of chain
M = 1000

# matrix that stores all posterior samples
post =matrix(0, M, 2)
dimnames(post)[[2]]= c("mu", "sigma2")

# initial values
mu = 4
sigma2=1

for (i in 1:M){
  mu     = post[i,1] = fc.mu(y, sigma2)
  sigma2 = post[i,2] = fc.sigma2(y, mu)
}

## Plot trace-plots & acf & density
par(mfrow=c(2,3))
for (i in 1:2) {
 plot(post[1:100,i],type="l")
 acf(post[1:100,i])
 hist(post[1:100,i], prob=T, col=0)
 lines(density(post[1:100,i]), col=2, lwd=4)
```

Example 3:

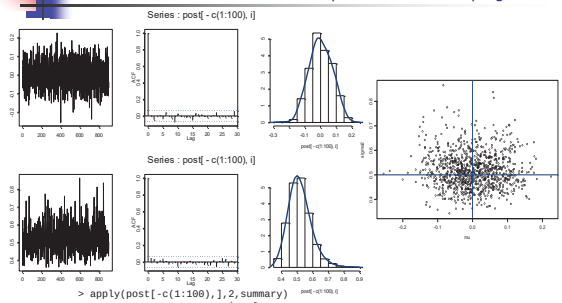
Normal model with noninformative priors & Gibbs Sampling



First 100 iterations…

Example 3:

Normal model with noninformative priors & Gibbs Sampling




```
> apply(post[-c(1:100),],2,summary)
                mu       sigma2
   Min. -0.254702905 0.3595476
1st Qu. -0.048932795 0.4702053 > apply(post[-c(1:100),],2, quantile, c(0.025,0.975))
 Median -0.001961042 0.5103874                mu    sigma2
   Mean  0.001535104 0.5181564   2.5% -0.1354625 0.3987192
3rd Qu.  0.054810665 0.5604048  97.5%  0.1423376 0.6888362
   Max.  0.225385852 0.8645269
```

Example 3:
Normal model with noninformative priors

- Prediction of a new observation
  - In this model, the posterior predictive distribution is a Student-t with (n-1) degrees of freedom, location at the sample mean, scale $s\sqrt{1+1/n}$
  - Analytical results not always available. What to do?
  - Recall

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

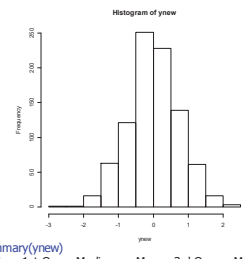
$$= \int P(Y_{NEW} \mid \theta) P(\theta \mid Data)\, d\theta$$

  - Can get samples from predictive distribution:
    - Draw samples Y_{NEW} given a posterior sample of the parameters

Example 3:
Normal model with noninformative priors

- Prediction of a new observation
  - Using simulation draws

```
post = post[-c(1:100),]
nsamples = nrow(post)
ynew = rnorm(nsamples, post[,1], sqrt(post[,2]))
hist(ynew)
```




```
> summary(ynew)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.750000 -0.451100 -0.006216  0.009239  0.477600  2.135000
```


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/MacOS 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

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

Table 6.2: Discrete univariate distributions in the **bugs** module

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

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

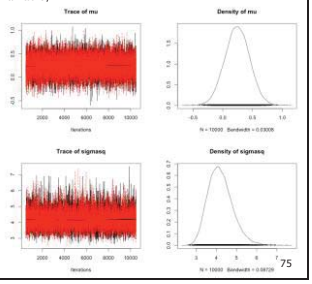
```
## define jags model within R
jags.m <- jags.model(file="example1.jag", data=data, inits=inits,
            n.chains=2, n.adapt=500)

## 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       0.005047

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 4.5740 5.5636
```



Convergence Diagnostics Methods

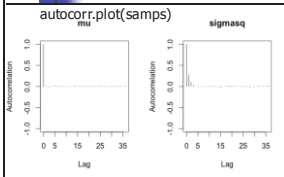
- Brooks, Gelman & Rubin
  - Two or more parallel chains (different starting values)
  - Comparison of within and between chain variance for each variable using the second half of chains
  - *"Rule-of-thumb": Samples are considered to arise from the stationary distribution if estimates are approximately 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

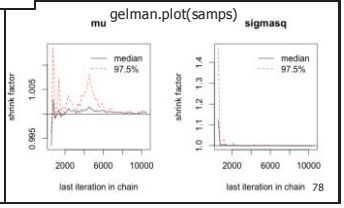
autocorr.plot(samps)



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]    -1.861522 0.0000 0.000000       0.000000
..
y[100]  -1.058556 0.0000 0.000000       0.000000
y[101]   0.297378 2.0084 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
```


Analyzing FEV Data Set with Jags

```
## Analysis of the FEV dataset using Jags
## read FEV data set
data = read.table("https://raw.githubusercontent.com/rhubb/STSCR2017/master/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])
  }
}
```


Analyzing 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 0.003083 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 0.89911
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
```

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY *of* WASHINGTON
School of Public Health

Meta-analysis

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, ScD, SM; Mary E. Wilson, MD; Elizabeth 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*

	Population		Cases of TB			TB Death		
Source, y	BCG	No BCG	BCG	No BCG	RR	BCG	No BCG	RR
Aronson,[20] 1948†	123	139	4	11	0.41	0	4	0.14
Ferguson and Simes,[46] 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,[47] 1977	13 598	12 867	62	248	0.24	. . .	. . .	. . .
Frimodt-Moller et al,[48] 1973	5069	5808	33	47	0.80	. . .	. . .	. . .
Stein and Aronson,[49] 1953	1541	1451	180	372	0.46	. . .	. . .	. . .
Vandiviere et al,[50] 1973	2545	629	8	10	0.20	. . .	. . .	. . .
Madras,[51] 1980§	88 391	88 391	505	499	1.01	. . .	. . .	. . .
Coetzee and Berjak,[52] 1968‖	7499	7277	29	45	0.63	. . .	. . .	. . .
Rosenthal et al,[43] 1961¶	1716	1665	17	65	0.25	1	6	0.16
Comstock et al,[17] 1974	50 634	27 338	186	141	0.71	8	12	0.36
Comstock and Webster,[44] 1969#	2498	2341	5	3	1.56	. . .	. . .	. . .
Comstock et al,[45] 1976#	16 913	17 854	27	29	0.98	. . .	. . .	. . .
Aronson et al,[11] 1958**	1541	1451	. . .	. . .	. . .	13	68	0.18
Levine and Sackett,[53] 1948††	566	528	. . .	. . .	. . .	6	6	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.[49]
††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

Data: $Y_i = \mu_i + \sigma_i \epsilon_i, i = 1, \ldots, k;\ \epsilon_i \sim N(0,1)$

Pooled Effect: $\bar{Y} = \frac{\sum_i W_i Y_i}{\sum_i W_i}$,

Fixed-effects Model (Mantel-Haenszel):

$$W_i = \frac{1}{V_i^2}$$

Random-effects Model (DerSimonian-Laird):
Assume $\mu_i \sim N(\mu, \tau^2)$

$$W_i = \frac{1}{V_i^2 + \widehat{\tau^2}}$$

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; Elizabeth 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)

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

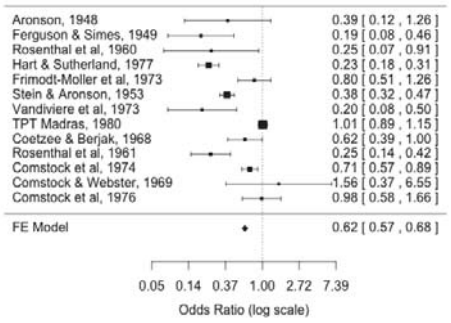
BCG Example

```
## 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



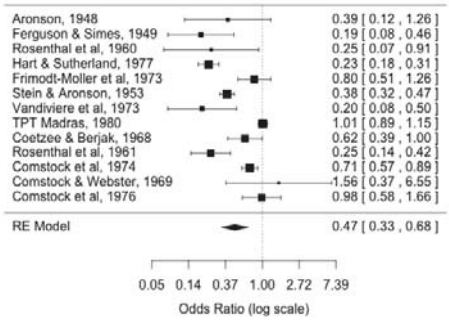
BCG Example (A): Standard Meta-Analysis
DerSimonian-Laird

```
> 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 as moderator
res.mr1 <- rma(yi, vi, mods = ~ ablat, data=dat)
res.mr1
```


BCG Example: 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
```

Some evidence that latitude is associated with observed effect size.

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 24  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  5888  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):

$$\text{logit}(p_i) = \delta_i + \beta_1 Group,$$

$$\delta_j \sim N(0, v_j^2), j = 1,\ldots,13;\ \beta_1 \sim N(0, \sigma_1^2)$$

```
> fit1 = inla(Y~ -1 + factor(trial) + factor(group), data=dat, Ntrials=N, family="binomial")
> summary(fit1)

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  -2.6817 0.2667     -3.1576  -2.5905     -2.1888 -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  -3.2183     -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     -5.3852  -5.1859     -5.0785 -5.1849   0
factor(trial)12 -6.1843 0.3541     -6.9374  -6.1635     -5.5445 -6.1281   0
factor(trial)13 -6.2258 0.1346     -6.4986  -6.2217     -5.9695 -6.2158   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 2 (B):

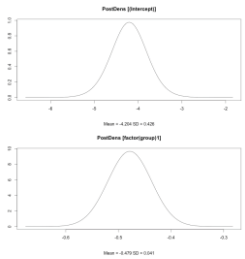
$$\text{logit}(p_i) = \delta_i + \beta_1 Group,$$
$$\delta_j \sim N(0, v^2), j = 1,\ldots,13;$$
$$v^2 \sim IG(a,b);\ \beta_1 \sim N(0, \sigma_1^2)$$

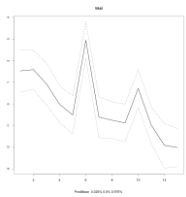
```
> fit2 = inla(Y~ factor(group) + f(trial, model="iid", param=c(0.001,0.001)), data=dat, Ntrials=N, family="binomial")
> summary(fit2)

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    -3.3660 -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:  -299.55
```

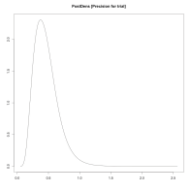
The overall posterior median OR=exp(-0.48)=0.62 (95% PCI= 0.57,0.67)
Posterior median precision = 0.43 (posterior median variance = 1/.43=2.33)

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

Model 2 (B):







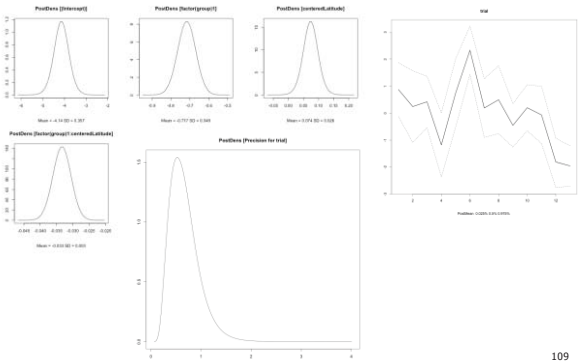
Model 3 (B):

$$\text{logit}(p_i) = \delta_i + \beta_1 Group + \beta_2 Latitude + \beta_3 Group * Latitude,$$
$$\delta_j \sim N(0, v^2), j = 1,\ldots,13;\ v^2 \sim IG(a,b);$$
$$\beta_k \sim N(0, \sigma_k^2); k = 1,2,3$$

```
> summary(fit3)

Time used:
 Pre-processing    Running inla Post-processing           Total
         0.0933          0.0473          0.0649          0.2955

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 3 (B):



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

Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons

CASE STUDY

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)***

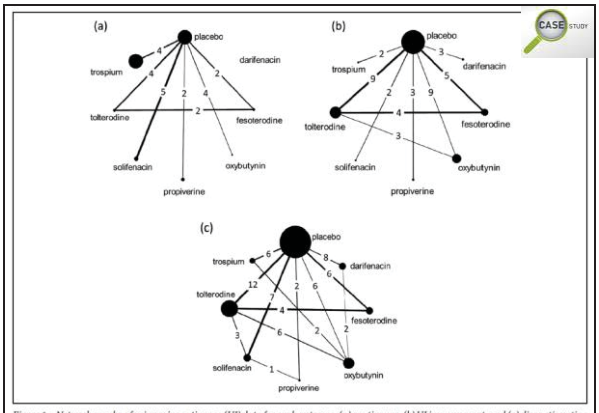


Figure 1 Network graphs of urinary incontinence (UI) data for each outcome: (a) continence, (b) UI improvement, and (c) discontinuation due to adverse events (AEs). The size of each node represents the number of studies investigating the drug, and the thickness of each edge implies the total number of 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("https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/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
 preA <- pow(.38,-2)
 # external info on A.
 A ~ dnorm(-2.6,preA)

 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
 preA <- pow(.38,-2)
 # external info on A.
 A ~ dnorm(-2.6,preA)
 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]    0.0736 0.0260 0.000153      0.000153
T[2]    0.0982 0.0334 0.000263      0.000288
T[3]    0.1432 0.0475 0.000348      0.000339
T[4]    0.1540 0.0547 0.000369      0.000526
best[1] 0.0000 0.0000 0.000000      0.000000
best[2] 0.0000 0.0000 0.000000      0.000000
best[3] 0.3150 0.4648 0.003272      0.003026
best[4] 0.6540 0.4648 0.003272      0.003026
d[1]    0.0000 0.0000 0.000000      0.000000
d[2]    0.2228 0.1269 0.000823      0.002422
d[3]    0.7652 0.0574 0.000490      0.001966
d[4]    0.8494 0.1741 0.001233      0.000385
lor[1,1] 0.0000 0.0000 0.000000      0.000000
lor[2,1] -0.2228 0.1269 0.000823      0.002422
lor[3,1] -0.7652 0.0574 0.000490      0.001966
lor[4,1] -0.8494 0.1741 0.001233      0.000385
lor[1,2] 0.2228 0.1269 0.000823      0.002422
lor[2,2] 0.0000 0.0000 0.000000      0.000000
lor[3,2] -0.5370 0.1345 0.000836      0.002570
lor[4,2] -0.6266 0.1935 0.001372      0.004221
lor[1,3] 0.7652 0.0574 0.000490      0.001966
lor[2,3] 0.5370 0.1345 0.000836      0.002570
lor[3,3] 0.0000 0.0000 0.000000      0.000000
lor[4,3] -0.0822 0.1734 0.001258      0.007751
lor[1,4] 0.8494 0.1741 0.001233      0.000385
lor[2,4] 0.6266 0.1935 0.001372      0.004221
lor[3,4] 0.0822 0.1734 0.001258      0.007751
lor[4,4] 0.0000 0.0000 0.000000      0.000000
rk[1]   3.9620 0.1826 0.001292      0.002580
rk[2]   3.0380 0.1985 0.001302      0.002454
rk[3]   1.6840 0.4648 0.003272      0.003026
rk[4]   1.3180 0.4648 0.003200      0.003417
```

```
> 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]    0.0736 0.0260 0.000260      0.000260
T[2]    0.1226 0.0580 0.000428      0.000149
T[3]    0.1524 0.0526 0.000420      0.000574
T[4]    0.2184 0.0976 0.000777      0.002472
best[1] 0.0000 0.0000 0.000000      0.000000
best[2] 0.6495 0.2181 0.001531      0.006512
best[3] 0.1835 0.3920 0.002591      0.003136
best[4] 0.7529 0.4253 0.003851      0.006888
d[1]    0.0000 0.0000 0.000000      0.000000
d[2]    0.5329 0.3958 0.002944      0.005176
d[3]    0.8124 0.2387 0.001728      0.006416
d[4]    1.1839 0.4566 0.003235      0.003618
lor[1,1] 0.0000 0.0000 0.000000      0.000000
lor[2,1] -0.5329 0.3958 0.002944      0.005176
lor[3,1] -0.8124 0.2387 0.001728      0.006416
lor[4,1] -1.1839 0.4566 0.003235      0.003618
lor[1,2] 0.5329 0.3958 0.002944      0.005176
lor[2,2] 0.0000 0.0000 0.000000      0.000000
lor[3,2] -0.2795 0.4015 0.002866      0.006623
lor[4,2] -0.6510 0.4383 0.003483      0.005579
lor[1,3] 0.8124 0.2387 0.001728      0.006416
lor[2,3] 0.2795 0.4015 0.002866      0.006623
lor[3,3] 0.0000 0.0000 0.000000      0.000000
lor[4,3] -0.3636 0.4582 0.003267      0.006984
lor[1,4] 1.1839 0.4566 0.003235      0.003618
lor[2,4] 0.6510 0.4383 0.003483      0.005579
lor[3,4] 0.3636 0.4582 0.003267      0.006984
lor[4,4] 0.0000 0.0000 0.000000      0.000000
rk[1]   3.9180 0.2934 0.002728      0.007927
rk[2]   2.7690 0.6548 0.004921      0.007924
rk[3]   2.0210 0.6280 0.004468      0.005797
rk[4]   1.3080 0.5881 0.004235      0.006658
sd      0.8387 0.1828 0.001243      0.007195
```


Meta-Analysis:
Mixed and Indirect Treatment Comparisons

```
(model1)

2. Quantiles for each variable:

          2.5%     25%     50%     75%   97.5%
T[1]    0.0238 0.0560 0.0731 0.0911 0.1260
T[2]    0.0418 0.0897 0.0826 0.1056 0.1894
T[3]    0.0676 0.1012 0.1395 0.1753 0.2560
T[4]    0.0695 0.1241 0.1475 0.1853 0.2899
best[1] 0.0000 0.0000 0.0000 0.0000 0.0000
best[2] 0.0000 0.0000 0.0000 0.0000 0.0000
best[3] 0.0000 0.0000 0.0000 1.0000 1.0000
best[4] 0.0000 0.0000 1.0000 1.0000 1.0000
d[1]    0.0000 0.0000 0.0000 0.0000 0.0000
d[2]    0.0249 0.1417 0.2268 0.3171 0.4287
d[3]    0.6285 0.7284 0.7852 0.8362 0.8714
d[4]    0.5093 0.7874 0.8415 0.9987 1.1863
lor[1,1] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[2,1] -0.4787 -0.3171 -0.2268 -0.1417 -0.0249
lor[3,1] -0.8714 -0.8362 -0.7852 -0.7284 -0.6285
lor[4,1] -1.1863 -0.9987 -0.8415 -0.7874 -0.5093
lor[1,2] -0.0249 0.1417 0.2268 0.3171 0.4287
lor[2,2] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[3,2] -0.8040 -0.6254 -0.5365 -0.4447 -0.2571
lor[4,2] -1.0000 -0.7652 -0.6266 -0.4852 -0.2873
lor[1,3] 0.6285 0.7284 0.7852 0.8362 0.8714
lor[2,3] 0.2571 0.4447 0.5365 0.6254 0.8040
lor[3,3] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[4,3] -0.4189 -0.1292 -0.0815 0.0212 0.2543
lor[1,4] 0.5093 0.7874 0.8415 0.9987 1.1863
lor[2,4] 0.2873 0.4852 0.6266 0.7652 1.0000
lor[3,4] -0.2543 -0.0212 0.0815 0.1292 0.4189
lor[4,4] 0.0000 0.0000 0.0000 0.0000 0.0000
rk[1]   3.0000 4.0000 4.0000 4.0000 4.0000
rk[2]   1.0000 2.0000 3.0000 3.0000 4.0000
rk[3]   1.0000 1.0000 2.0000 2.0000 2.0000
rk[4]   1.0000 1.0000 1.0000 2.0000 2.0000
```

```
(model2)

2. Quantiles for each variable:

          2.5%     25%     50%     75%   97.5%
T[1]    0.0240 0.0520 0.0877 0.0944 0.1357
T[2]    0.0456 0.0841 0.1183 0.1566 0.2860
T[3]    0.0525 0.1129 0.1450 0.1868 0.2510
T[4]    0.0696 0.1355 0.1832 0.2863 0.4413
best[1] 0.0000 0.0000 0.0000 0.0000 0.0000
best[2] 0.0000 0.0000 0.0000 0.0000 1.0000
best[3] 0.0000 0.0000 0.0000 0.0000 1.0000
best[4] 0.0000 1.0000 1.0000 1.0000 1.0000
d[1]    0.0000 0.0000 0.0000 0.0000 0.0000
d[2]    -0.2488 0.2805 0.5236 0.7945 1.3323
d[3]    0.3695 0.6580 0.8885 0.9789 1.3889
d[4]    0.2834 0.8965 1.1871 1.4788 2.1368
lor[1,1] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[2,1] -1.3328 -0.7945 -0.5236 -0.2805 0.2489
lor[3,1] -1.3986 -0.9789 -0.8885 -0.6580 -0.3695
lor[4,1] -2.1368 -1.4788 -1.1871 -0.8965 -0.2832
lor[1,2] -0.2488 0.2805 0.5236 0.7945 1.3323
lor[2,2] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[3,2] -1.0882 -0.5491 -0.2894 -0.0266 0.5945
lor[4,2] -1.6122 -0.9223 -0.6510 -0.3812 0.3888
lor[1,3] 0.3695 0.6580 0.8885 0.9789 1.3889
lor[2,3] -0.5946 0.0266 0.2894 0.5491 1.0882
lor[3,3] 0.0000 0.0000 0.0000 0.0000 0.0000
lor[4,3] -1.2887 -0.6763 -0.3812 -0.0266 0.5276
lor[1,4] 0.2834 0.8965 1.1871 1.4788 2.1368
lor[2,4] -0.3888 0.3348 0.6510 0.9223 1.6122
lor[3,4] -0.5276 0.0266 0.3812 0.6763 1.2887
lor[4,4] 0.0000 0.0000 0.0000 0.0000 0.0000
rk[1]   3.0000 4.0000 4.0000 4.0000 4.0000
rk[2]   1.0000 2.0000 3.0000 3.0000 4.0000
rk[3]   1.0000 2.0000 2.0000 2.0000 3.0000
rk[4]   1.0000 1.0000 1.0000 1.0000 3.0000
sd      0.5491 0.6943 0.8851 0.9327 1.2588
```

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Diagnostic testing with missing gold standard

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

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	0.78	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.

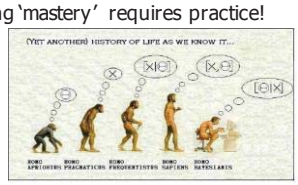
Reproducing analyses: Using only one diagnostic test

```
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!