



## Bayesian Methods for Clinical Research: Computational Methods and Application

1

---

---

---

---

---

---

### Outline

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

2

---

---

---

---

---

---

### Overview of the Bayesian approach

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

3

---

---

---

---

---

---

## The likelihood function

- A **likelihood function** is a function of the parameters of a probability model given the outcomes.
  - The *likelihood* of  $\theta$ , given outcome  $y$ , is equal to the *probability* of that observed outcome given  $\theta$ .
- 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  $\theta$  is:
      - $L(\theta|y=1) = P(Y=1|\theta)=\theta$
    - Given that  $y=0$ , the likelihood function of  $\theta$  is:
      - $L(\theta|y=0) = P(Y=0|\theta)=1-\theta$

4

---

---

---

---

---

## 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 is crucial in assessing 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

5

---

---

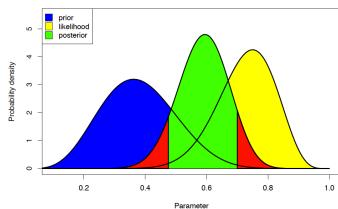
---

---

---

## Posterior distribution

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



6

---

---

---

---

---



## Inference based on posterior distribution

- Inferences based on summaries of the posterior distribution

- Point estimates:

- Mean/Median/Mode

- Interval estimates:

- One-sided credible intervals

- Two-sided credible intervals

- Equi-tail area

- Narrowest interval

[HPD: highest posterior density intervals]

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

7

---

---

---

---

---

---

---



## Primary packages we will use



- INLA

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

```
>install.packages("INLA", repos="https://www.math.ntnu.no/inla/R/stable")
>library(INLA)
```

- rjags (alternative choices R2jags, runjags)

8

---

---

---

---

---

---

---



DEPARTMENT OF BIOSTATISTICS

UNIVERSITY of WASHINGTON

School of Public Health



## Bayesian Regression Models

9

---

---

---

---

---

---

---

Generalized Linear Regression Models

- Mean:  $E[Y_i | X_{i1}, X_{i2}, \dots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$  where  $g$  is a link function
  - Regression Model:  $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$ 
    - Linear regression model  

$$g(\mu_i) = \mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$
    - Logistic regression model  

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$
    - Probit regression model  

$$g(\mu_i) = \Phi^{-1}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$
    - Poisson regression model  

$$g(\mu_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$

10

Bayesian GLM

- Mean:  $E[Y_i | X_{i1}, \dots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$  where  $g$  is a link function
  - Regression Model:  $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$
  - Priors:**
    - Regression parameters:  $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$
    - "Nuisance" parameters (e.g. in linear regression  $\sigma^2$ )
  - 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)

11

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

12

## Bayesian GLM in R: INLA package

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

- Regression Model:

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

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

$\beta_j$ : linear effects of covariates  $X$

$\varepsilon_i$ : unstructured terms

- Assumption in latent Gaussian models:

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

13

## Bayesian GLM in R: INLA package

- Latent model:

Let  $z$  represent the collection of all Gaussian variables:

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

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

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

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

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

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

14

Latent Models in INLA	Model	Name	Description
For more info: <a href="http://www.r-inla.org">http://www.r-inla.org</a>	Independent random variables	iid	indep.pdf
	Linear	linear	linear.pdf
	Constrained linear	clinear	clinear.pdf
	Random walk of order 1	rw1	rw1.pdf
	Random walk of order 2	rw2	rw2.pdf
	Continuous random walk of order 2	crw2	crw2.pdf
	Model for seasonal variation	seasonal	seasonal.pdf
	Model for spatial effect	bessag	bessag.pdf
	Model for spatial effect	bessagproper	bessagproper.pdf
	Model for weighted spatial effects	bessag2	bessag2.pdf
	Model for spatial effect + random effect	bym	bym.pdf
	Autoregressive model of order 1	arl	arl.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
	Sigmoid and reverse sigmoidal	sigm, revsigm	sigm.pdf

Likelihoods

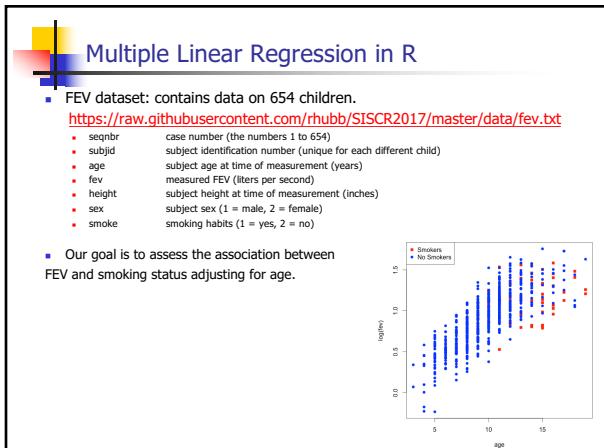
Negative Binomial	nbnnomial	nbnnomial.pdf
Poisson	poisson	poisson.pdf
Binomial	binomial	binomial.pdf
Gaussian	gaussian	gaussian.pdf
Skew Normal	sn	sn.pdf
Student-t	t	Student-t.pdf
Gaussian model for stochastic volatility	stochvol	stochvolgaussian.pdf
Student-t model for stochastic volatility	stochvol.t	stochvolt.pdf
NIG model for stochastic volatility	stochvol.nig	stochvolnig.pdf
Zero inflated Poisson	zeroinflated.poisson.0	zeroinflated.pdf
	zeroinflated.poisson.1	
	zeroinflated.poisson.2	
Zero inflated Binomial	zeroinflated.binomial.0	zeroinflated.pdf
	zeroinflated.binomial.1	
Zero inflated negative Binomial	zeroinflated.nbinomial.0	zeroinflated.pdf
	zeroinflated.nbinomial.1	
	zeroinflated.nbinomial.2	
Zero inflated beta binomial (type 0/1)	zeroinflated.bbbinomial.0	zeroinflated.pdf
	zeroinflated.bbbinomial.1	
Zero inflated beta binomial (type 2)	zeroinflated.bbbinomial.2	zeroinflated.bbbin.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 mode)	weibull	weibull.pdf
Log-logistic (Survival model)	loglogistic	loglogistic.pdf
Log-normal (Survival model)	lognormal	lognormal.pdf
Cox model (Survival model)	coxph	coxph.pdf

16

Priors on hyperparameters

Model	Name	Description
Normal distribution	normal	gaussian.pdf
Log-gamma distribution	gaussian	prior-loggamma.pdf
Improper flat prior	flat	prior-flat.pdf
Truncated Normal distribution	logtnormal	log-tnormal.pdf
logtnormal	logtgaussian	
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	iid123.pdf
Beta for correlations	betacorrelation	betacorrelation.pdf
Logit of a Beta	logitbeta	logitbeta.pdf
Define your own prior	expression:	expression.pdf
Define your own prior	table:	table.pdf

17



## Multiple Linear Regression in R

```
> ## read FEV data set
> #url = "https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1892772/reader%20data/fev.txt"
> #colnames("fev.txt", header=1, sep=",", "height", "sex", "smoke")
> # read file
> # examine a few entries of the data set
> head(data)

#> # summarize the variables
> summary(data)

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

#> # scatter plot of log(fev) by sex, but stratified by smoking status
> points(log(fev) ~ age, col="red", pch=15, data=data)
> points(log(fev) ~ age, col="blue", pch=16, data=data)
> legend("topleft", c("Smokers", "No Smokers"), col=c("red", "blue"), pch=c(15,16))

> # summarize the variables
> summary(data)
```

19

## 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        0.08992701 0.030091877 0.03083392 0.08992617 0.148969457 0.08992701 4.070901e-14
age          0.09076807 0.003050123 0.08477837 0.09076798 0.096752630 0.09076807 4.661605e-14

> fit$summary.hyp
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
```

20

## 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(
    prior = list(
      prior = "normal",
      param = c(0, 10)
    )
  )
)
fit.prior1$summary.fix
fit.prior1$summary.hyp
```

Making prior assumptions explicit

```
> fit.prior1$summary.fix
mean      sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) -0.15691453 0.07908793 -0.31222632 -0.15691674 -0.001736104 -0.15691451 4.714742e-14
smoke        0.08992701 0.031641908 0.02778911 0.08992612 0.152011540 0.08992701 4.721406e-14
age          0.09076807 0.003207235 0.08446975 0.09076798 0.097060977 0.09076807 4.216079e-14
> fit.prior1$summary.hyp
mean      sd 0.025quant 0.5quant 0.975quant mode
Precision for the Gaussian observations 20.42931 1.169583 18.20924 20.403 22.89734 20.35767
```

21

## 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:  $H(t)$

22

---



---



---



---



---



---



---



---

## 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:  
 $[h(t)]^d S(t)$

23

---



---



---



---



---



---



---



---

## Survival Models: Proportional Hazards

- Proportional Hazards (PH) Model:

$$h(t) = h_0(t) \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)$$

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

24

---



---



---



---



---



---



---



---

## 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
  progresc 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?

25

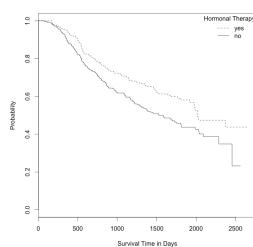
## PH regression models in R

```

## -- data publicly available in an R-package
## library(survival, package="TH.data")
## summary(GBSG2)

```

	age	tsize	tgrade	pnodes	progresc	estrec	time	cens
Min.	21.00	3.00	I : 81	Min. : 1.00	0.00	Min. : 0.00	Min. : 8.0	0
1st Qu.	46.00	20.00	II : 1444	1st Qu.: 1.00	1.00	1st Qu.: 8.00	1st Qu.: 567.8	1
Median	50.00	21.00	III : 161	Median : 1.00	1.00	Median : 11.00	Median : 1114.5	1
Mean	53.05	29.33		Mean : 5.01	Mean : 110.0	Mean : 96.25	Mean : 1124.5	1
3rd Qu.	61.00	35.00		3rd Qu.: 7.00	3rd Qu.: 131.8	3rd Qu.: 114.00	3rd Qu.: 1186.8	1
Max.	86.00	120.00		Max. :15.00	Max. :151.0	Max. :1544.0	Max. :12699.0	1



26

## 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  0.003 ** 
...
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 = 0.013 )
Esquare= 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=0.003423

```

```

## Parametric survival (Weibull regression)
> library(survival)
> 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  0.604     0       1   (reference)
  yes 0.396   -0.393  0.675  0.125  0.002
...
log(scale)        7.408     0.058  0.000
log(shape)        0.251     0.050  0.000

Events      299
Total time at risk      771400
Risks per unit time     2.000E-05
LR test statistic      10.36
Degrees of freedom       1
Overall p-value        0.000128745

```

## Bayesian PH regression models in R: Non-parametric

```

library(rms)
## Bayesian non-parametric PH model
# fit <- inla(inla.surv(time, cens) ~ horTh, family="coxph", data=GSG2)
# summary(fit)

Call:
Pre-processing    Running inla Post-processing      Total
  0.0946          0.5117          0.0574          0.6639

Fixed effects:
mean      sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) -7.7078 0.1403   -7.9948  -7.7039  -7.4426  -7.6965  0
horTh     -0.3860 0.1249   -0.6145  -0.3650  -0.1237  -0.3628  0

Random effects:
baseline.hazard  RML model

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

Expected number of effective parameters (std dev): 9.484(1.09)
Number of equivalent replicates: 497.48
Marginal Likelihood: -1379.80

```

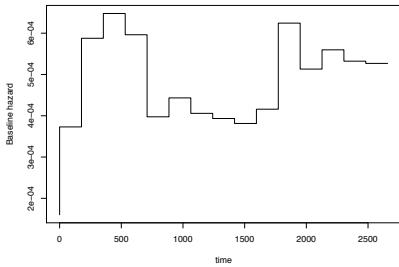
28

## Bayesian PH regression models in R: Non-parametric

```

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

```



29

## Bayesian PH regression models in R: Parametric

```

## Bayesian parametric PH model
# fit <- inla.surv(time, cens) ~ horTh, family = "weibull", data = GSG2)
# summary(fit)

Call:
#(*inla(formula = inla.surv(time, cens) ~ horTh, family = 'weibull', *, * data = GSG2))

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  -8.8047  -9.3598  1e+04
horTh     -0.3891 0.1248   -0.6373  -0.3880  -0.1470  -0.3859  0e+00

The model has no random effects

Model hyperparameters:
alpha parameter for weibull 1.2651 0.0749 1.1438 1.2557 1.4339 1.1229
Expected number of effective parameters (std dev): 2.005(0.00)
Number of equivalent replicates: 342.15
Marginal Likelihood: -2641.95

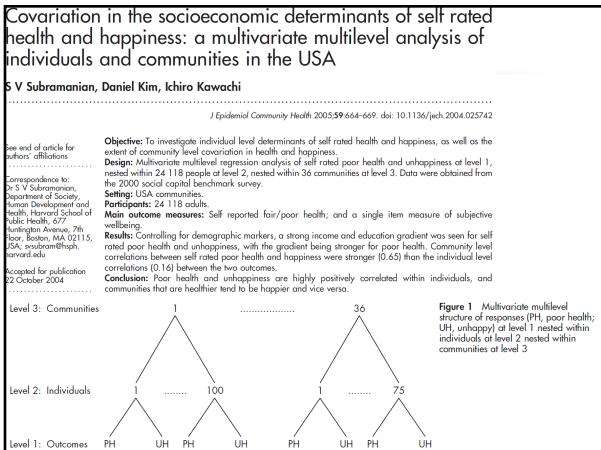
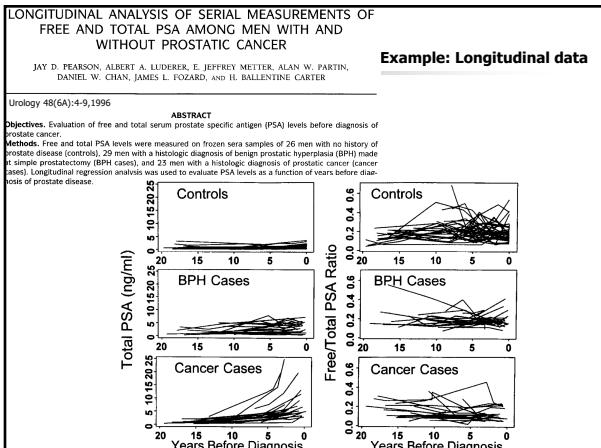
```

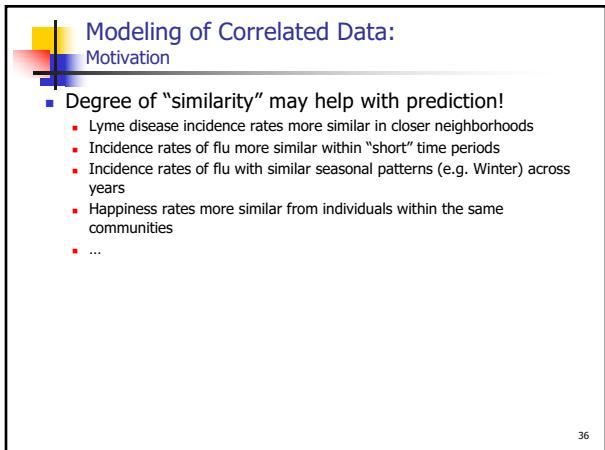
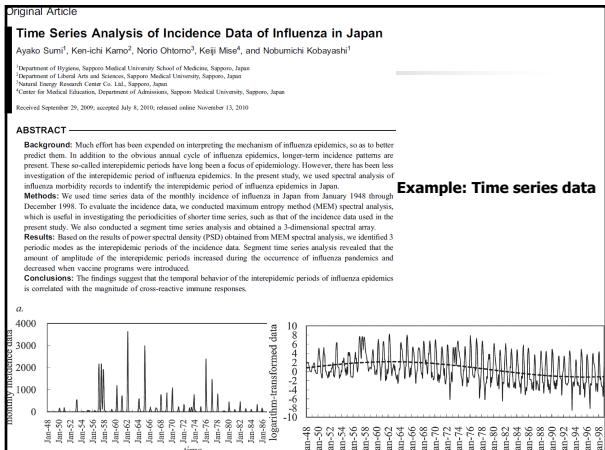
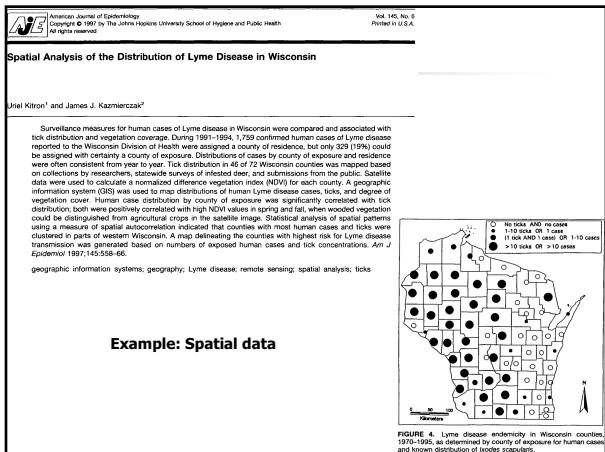
30



## Analysis of Correlated Data

31

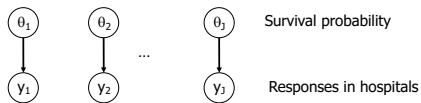




## Hierarchical Model Example:

- Goal:

- Study the effectiveness of cardiac treatments



Independent Data  
(Separate analysis using data from each study)

37

---

---

---

---

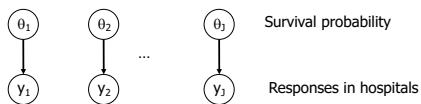
---

---

## Hierarchical Model Example:

- Goal:

- Study the effectiveness of cardiac treatments



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

38

---

---

---

---

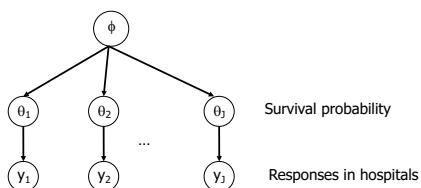
---

---

## Hierarchical Model Example:

- Goal:

- Study the effectiveness of cardiac treatments



This implies, marginally, correlation between observations!

39

---

---

---

---

---

---

## Hierarchical Model Example:

- Goal:

- Study the effectiveness of cardiac treatments
  - $\theta_j$ : survival probability for patients in hospital  $j$
  - $\phi$  : overall survival probability

- Inference:

- Estimate  $\theta_j$ 's borrowing strength of information from all other hospitals
  - Estimate  $\phi$  taking into account the variability among hospitals

40

---

---

---

---

---

---

---

## Hierarchical Model: Exchangeability

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

- Example:

$$\begin{aligned} P(Y_1, Y_2, Y_3) &= P(Y_2, Y_1, Y_3) = P(Y_2, Y_3, Y_1) = \\ &= P(Y_1, Y_3, Y_2) = P(Y_3, Y_1, Y_2) = \\ &= P(Y_3, Y_2, Y_1) \end{aligned}$$



we would judge  $Y_1, Y_2, Y_3$   
exchangeable!

41

---

---

---

---

---

---

---

## Hierarchical Model: Exchangeability

- Note:

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

42

---

---

---

---

---

---

---

**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

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

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

	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...

44

**Hierarchical Model:**  
Exchangeability: De Finetti's theorem

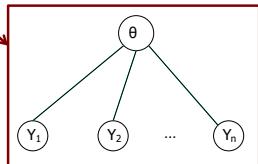
- For all infinite sequences of exchangeable random binary variables  $\{Y_1, Y_2, \dots\}$ , there corresponds a distribution function  $F$  on  $(0,1)$  such that for all  $n$  and  $k \leq n$ ,

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

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

45

## Hierarchical Model:



46

## Hierarchical Model:

Exercise (back to example)

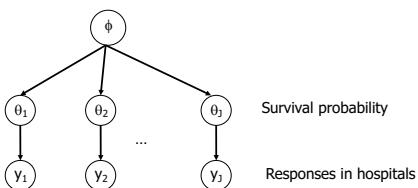
- Where would we assume exchangeability?

A hierarchical tree diagram. At the top is a circle labeled  $\phi$ . It has three arrows pointing down to three circles labeled  $\theta_1$ ,  $\theta_2$ , and  $\theta_j$ . Below  $\theta_1$  is a circle labeled  $y_1$ . Below  $\theta_2$  is a circle labeled  $y_2$ . Between  $\theta_2$  and  $\theta_j$  is an ellipsis (...). Below  $\theta_j$  is a circle labeled  $y_j$ .

Survival probability

Responses in hospitals

47



47



## 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
$$\pi_1(\theta|0_1), \pi_2(\theta_1|\theta_2), \dots, \pi_n(\theta_{n-1}|\theta_n)$$
and a marginal distribution  $\pi_{n+1}(\theta_n)$  such that
$$\pi(\theta) = \prod \pi_1(\theta|0_1), \pi_2(\theta_1|\theta_2), \dots, \pi_n(\theta_{n-1}|\theta_n) \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n$$
- Parameters  $\theta_i$  are called hyperparameters 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
  - Sensitivity analysis is very important in hierarchical modeling

48

## 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?
    - ...

49

---

---

---

---

---

---

---

## 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/rhub/SISC2017/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

50

---

---

---

---

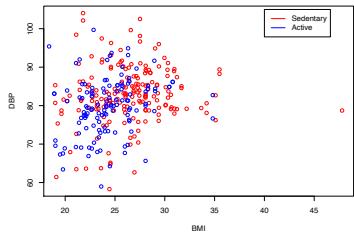
---

---

---

## 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



51

---

---

---

---

---

---

---

**Hierarchical Model for HAREST study**

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

```

graph TD
    phi((phi)) --> theta1((theta_1))
    phi --> theta2((theta_2))
    phi --> thenn((...))
    phi --> thennn((theta_n))
    theta1 --> y10((y_{1,0}))
    theta1 --> y11((y_{1,1}))
    theta1 --> y12((y_{1,2}))
    theta2 --> y20((y_{2,0}))
    theta2 --> y21((y_{2,1}))
    theta2 --> y22((y_{2,2}))
    thenn --> yn0((y_{n,0}))
    thenn --> yn1((y_{n,1}))
    thenn --> yn2((y_{n,2}))
    subgraph "Population mean DBP"
        phi
        theta1
        theta2
        thenn
        thennn
    end
    subgraph "Mean DBP for patient k"
        theta1
        theta2
        thenn
        thennn
        y10
        y11
        y12
        y20
        y21
        y22
        yn0
        yn1
        yn2
    end
    subgraph "DBP for patient k at time t"
        y10
        y11
        y12
        y20
        y21
        y22
        yn0
        yn1
        yn2
    end

```

52

**Motivating example: HARVEST study**

```

> harvest <- read.csv("https://raw.githubusercontent.com/rhubb/S1SCR2017/master/data/harvest.csv",
  header = T, na.strings = ".")
> summary(harvest[,c("SBP","DBP","Age","BMI","Male")])

```

	SBP	DBP	Age	BMI	Male
Min.	:100.1	:48.23	:16.00	:14.53	:0.000
1st Qu.	:123.9	:77.92	:1st Qu.:26.00	:1st Qu.:22.85	:1st Qu.:0.000
Median	:130.2	:81.97	:Median :33.00	:Median :24.87	:Median :1.000
Mean	:131.1	:81.49	:Mean :33.24	:Mean :25.36	:Mean :0.729
3rd Qu.	:137.8	: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	

Number of participants

1 2 3

53

**Motivating example: HARVEST study**

```

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

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

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

Fixed effects:
            mean      sd 0.025quant 0.5quant 0.975quant mode
(Intercept) 79.0616 2.8079   73.5402 79.0625   84.5726 79.0644  0
factor(Sport)1 -1.9382 1.4124  -4.7531 -1.9725   0.7895 -1.9705  0
factor(Sport)2 -1.3390 1.4557  -3.1579 -1.3602   0.8180 -1.3589  0
factor(Sport)3 -4.4832 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

Model hyperparameters:
            mean      sd 0.025quant 0.5quant 0.975quant mode
Precision for the Gaussian observations 0.0421 0.0053  0.0322 0.0419  0.0529 0.0418
Precision for ID                         0.0255 0.0039  0.0190 0.0251  0.0342 0.0243

```

Precision of DBP across participants

54



## Markov Chain Monte Carlo (MCMC)

(Implementation via JAGS)

55

---

---

---

---

---

---

## Markov Chains

### Definition:

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

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

### Definition:

- A Markov Chain is homogeneous if

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



56

---

---

---

---

---

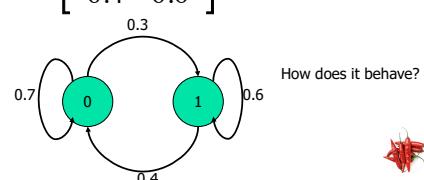
---

## Markov Chains

### Example:

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

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



57

---

---

---

---

---

---

## Markov Chains

- Transition matrix in n steps?

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

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

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

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



58

## Markov Chains

- Limiting distribution:

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

- Note that:

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

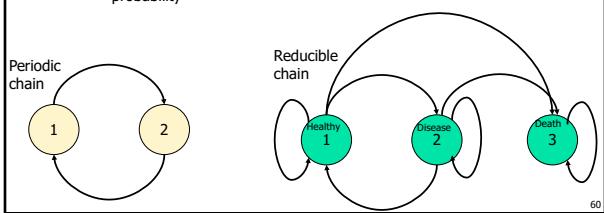


59

## Markov Chains

- Conditions for convergence:

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



60

## Markov Chains and MCMC

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

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

61

## MCMC algorithms

- 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):  $[\theta_i | \theta_{-i}]$
- Algorithm:
  1. Set starting values  $(\theta_1^0, \dots, \theta_k^0)$
  2. Iteration i
    - a. Draw  $\theta_1^{(i)} \sim [\theta_1 | \theta_2^{(i-1)}, \dots, \theta_k^{(i-1)}]$
    - b. Draw  $\theta_2^{(i)} \sim [\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}, \dots, \theta_k^{(i-1)}]$
    - c. ...
    - d. Draw  $\theta_k^{(i)} \sim [\theta_k | \theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{k-1}^{(i)}]$
  3. Repeat step 1. After t iterations obtain

- Theorem:  
$$(\theta_1^{(t)}, \dots, \theta_k^{(t)}) \xrightarrow{D} [\theta_1, \dots, \theta_k] \text{ as } t \rightarrow \infty$$



63

## 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...

64

---

---

---

---

---

---

---

---

## Example 3:

### Normal model with noninformative priors

Let  $Y_i \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$  and  $\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ .

We had:

$$\begin{aligned}\pi(\mu, \sigma^2 | y) &\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \\ &\times \exp\left\{-\frac{\sum(y_i - \mu)^2}{2\sigma^2}\right\}\end{aligned}$$

Let  $\tau = 1/\sigma^2$ . Easy to derive:

$$\begin{aligned}\pi(\mu | \sigma^2, y) &= N(\bar{y}, \sigma^2/n) \\ \pi(\tau | \mu, y) &= \Gamma\left(\frac{n}{2}, \frac{1}{2} \sum (y_i - \mu)^2\right)\end{aligned}$$

65

---

---

---

---

---

---

---

---

## 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)
}
```

---

---

---

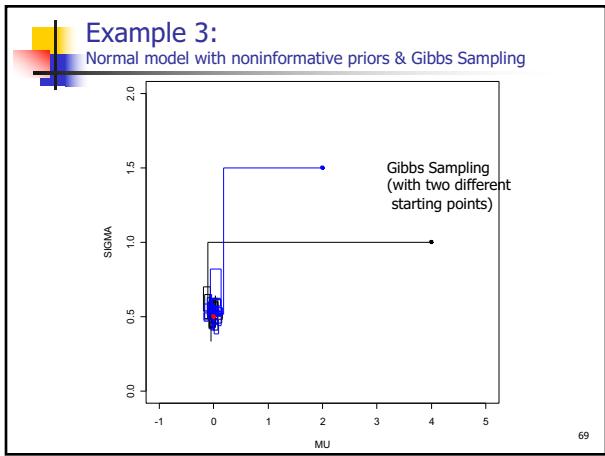
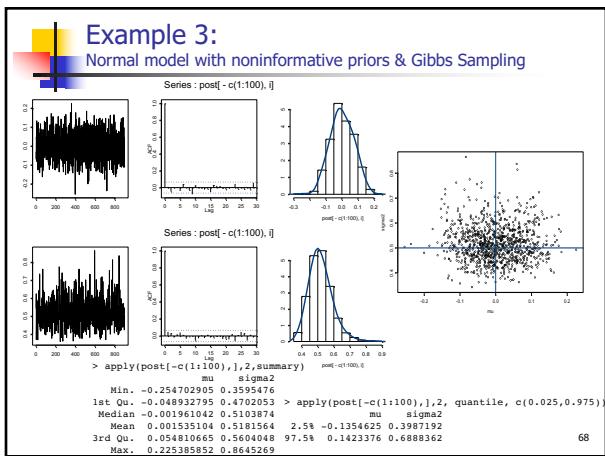
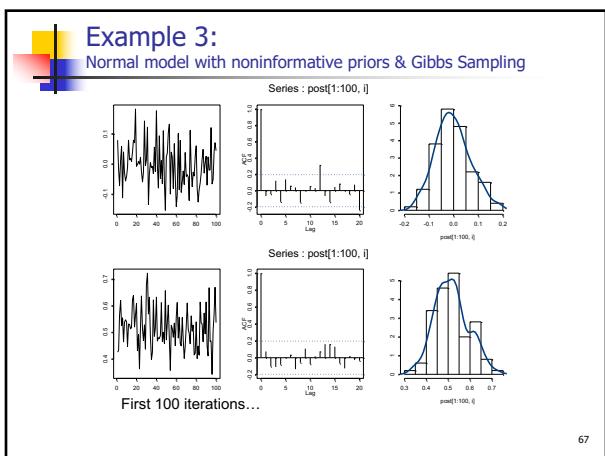
---

---

---

---

---



### Example 3:

Normal model with noninformative priors

- 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  $s\sqrt{1+\frac{1}{n}}$
- Analytical results not always available. What to do?
- Recall

$$\begin{aligned} P(Y_{\text{NEW}} | \text{Data}) &= \int P(Y_{\text{NEW}} | \text{Data}, \theta)P(\theta | \text{Data})d\theta \\ &= \int P(Y_{\text{NEW}} | \theta)P(\theta | \text{Data})d\theta \end{aligned}$$

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

70

---

---

---

---

---

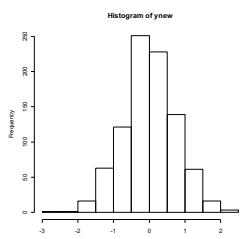
---

### 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
```

71

---

---

---

---

---

---

### 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

72

---

---

---

---

---

---

**Jags (Just Another Gibbs Sampler)**

- Website: <http://mcmc-jags.sourceforge.net>
- For MAC: <http://sourceforge.net/projects/mcmc-jags/files/JAGS/3.x/Mac%20OS%20X/>
- 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

73

**Using Jags**

Name	Usage	Density	Lower	Upper
Beta	<code>dbeta(a, b)</code>	$\frac{x^{a-1}(1-x)^{b-1}}{\beta(a, b)}$	0	1
Chi-square	<code>dchisq(k)</code>	$\frac{x^{k/2-1} \exp(-x/2)}{2^k \Gamma(\frac{k}{2})}$	0	
Double exponential	<code>ddexp(mu, tau)</code>	$\tau \exp(-\tau x-\mu )/2$		
Exponential	<code>dexp(lambda)</code>	$\lambda \exp(-\lambda x)$	0	
F	<code>df(x, m, n)</code>	$\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{x}{n+m}\right)^{\frac{m}{2}} x^{m/2-1} \left(1 + \frac{x}{n+m}\right)^{-\frac{n+m}{2}}$	0	
Gamma	<code>dgamma(r, lambda)</code>	$\lambda^r x^{r-1} \exp(-\lambda x)$	0	
Generalized gamma	<code>dggamma(r, lambda, b)</code>	$b \lambda^r x^b \exp(-(\lambda x)^b)$	0	
gamma	<code>dgamma(mu, r)</code>	$\frac{1}{\Gamma(r)} x^{r-1} \exp(-\mu x)$	0	
Logistic	<code>dlogis(mu, tau)</code>	$\frac{\tau \exp(x-\mu)\tau}{[1+\exp(x-\mu)]^2}$		
Log-normal	<code>dlnorm(mu, tau)</code>	$(\frac{x}{\tau})^{-1} \tau^{-1} \exp\{-\tau(\log(x)-\mu)^2/2\}$	0	
Noncentral chi-square	<code>dnchisq(k, delta)</code>	$\sum_{r=0}^{\infty} \frac{\exp(-\delta/2) (\frac{\delta}{2})^{(k+2r)/2} \exp(-r\chi^2/2)}{r!}$	0	
Chi-square	<code>dchisq(k)</code>	$\frac{1}{2^k \Gamma(\frac{k}{2})} \left(\frac{\chi^2}{2}\right)^{\frac{k}{2}-1} \exp(-\chi^2/2)$	0	
Normal	<code>dnorm(mu, tau)</code>	$(\frac{x-\mu}{\tau})^{-\frac{1}{2}} \exp(-\tau(x-\mu)^2/2)$		
Pareto	<code>dpareto(x, c)</code>	$c x^{-c-1}$	c	
Student t	<code>dt(mu, tau, k)</code>	$\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(\frac{\chi^2}{k\tau^2}\right)^{\frac{k-1}{2}} \left\{1 + \frac{\tau(\chi^2+k\mu^2)}{k}\right\}^{-\frac{(k+1)}{2}}$	a	b
Uniform	<code>dunif(a, b)</code>	$\frac{1}{b-a}$	a	b
Weibull	<code>dweib(v, lambda)</code>	$v \lambda x^{v-1} \exp(-\lambda x^v)$	0	

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

74

**Using Jags**

Name	Usage	Density	Lower	Upper
Beta	<code>dbetabin(n, a, b)</code>	$\binom{n}{x} \binom{n-a}{x} \binom{n+b-a}{n-x}^{-1}$	0	n
binomial	<code>dbinom(n, p)</code>	$p^n (1-p)^{n-n}$	0	1
Bernoulli	<code>dbin(1, p)</code>	$p(1-p)^{1-x}$		
Binomial	<code>dbin(n, p)</code>	$\binom{n}{x} p^x (1-p)^{n-x}$	0	n
Categorical	<code>ddcat(x, N)</code>	$\sum_{i=1}^N \frac{x_i}{N}$	1	N
Noncentral hypergeometric	<code>dhypgeom(n1, n2, m1, psi)</code>	$\frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\sum \binom{n_1}{x_1} \binom{n_2}{x_2} \psi^x} \psi^x$	max(0, n <sub>1</sub> -m <sub>1</sub> )	min(n <sub>1</sub> , m <sub>1</sub> )
hypergeometric	<code>dhyper(n1, n2, m1)</code>	$\frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{n_1+n_2}{x_1+x_2}} \psi^x$		
Negative binomial	<code>dnegbin(r, p)</code>	$\binom{x+r-1}{x-1} p^r (1-p)^x$	0	
binomial	<code>dgeom(mu, r)</code>	$(1-p)^{r-1} p^r$		
Poisson	<code>dpois(mu, lambda)</code>	$\exp(-\lambda) \lambda^x / x!$	0	
	$\lambda > 0$			

Table 6.2: Discrete univariate distributions in the bugs module

75

**Using Jags**

Name	Usage	Density
Dirichlet	<code>p ~ ddirich(alpha)</code> $\alpha_j > 0$	$\frac{p^{\alpha_j-1}}{\Gamma(\sum_i \alpha_i) \prod_j \Gamma(\alpha_j)}$
Multivariate normal	<code>x ~ dmvn(mu, Omega)</code>	$\left(\frac{ Omega }{2\pi}\right)^{\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^T \Omega (x-\mu))$
Wishart	<code>Omega ~ dwish(R, k)</code>	$\frac{ Omega ^{(k-p)/2}  R ^{k/2} \exp(-Tr(R)/2)}{2^{kp/2}  T_p(k/2) }$
Multivariate Student t	<code>x ~ dmvt(mu, Omega, k)</code>	$\frac{\Gamma((k+p)/2)}{\Gamma(k/2)(p/2)} \frac{1}{2^p} \left\{1 + \frac{1}{k} (x-\mu)^T \Omega (x-\mu)\right\}^{-\frac{k+p}{2}}$
Multinomial	<code>x ~ dmult(pi, n)</code> $\sum_j x_j = n$	$n! \prod_j \frac{\pi_j^{x_j}}{x_j!}$

Table 6.3: Multivariate distributions in the bugs module



Link function	Description	Range	Inverse
<code>cloglog(y) &lt;- x</code>	Complementary log log	$0 < y < 1$	$y \leftarrow \text{icloglog}(x)$
<code>log(y) &lt;- x</code>	Log	$0 < y$	$y \leftarrow \exp(x)$
<code>logit(y) &lt;- x</code>	Logit	$0 < y < 1$	$y \leftarrow \text{ilogit}(x)$
<code>probit(y) &lt;- x</code>	Probit	$0 < y < 1$	$y \leftarrow \phi(x)$

Table 5.4: Link functions in the `bugs` module

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

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

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

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

76

## 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)

77

## 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(mu0=0, sigmasq=1)
```

78

## 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)

```

79

## Example 1: using jags

```

> jagsmodel <- 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
Initialising model
[*****] 100%
## specify parameters to be monitored
> params <- c("mu","sigmasq")
> ## run jags and save posterior samples
> posterior <- jags(data = data, parameters = params, n.itern=10000)
[*****] 100%
## summarize posterior samples
> summary(samps)

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

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

2. Quantiles for each variable:
      2.5%   25%   50%   75% 97.5%
mu   -0.1595 0.1037 0.2430 0.3812 0.6408

```

80

## Example 1: using jags

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

```
Iterations = 1000:10500  
Thinning interval = 1  
Number of chains = 2  
Sample size per chain = 9
```

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

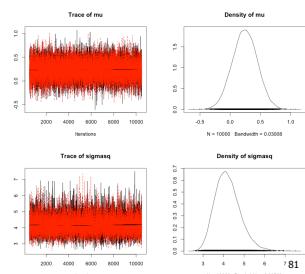
```

      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 3.1472 3.7230 4.1526 4.5748 5.5636

```



81

**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

---



---



---



---



---



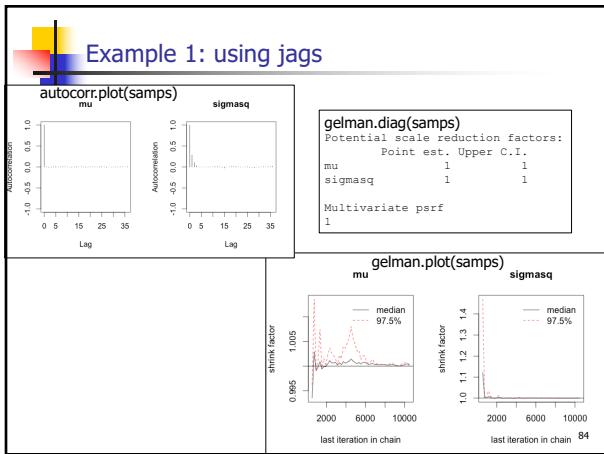
---



---



---



## Example 1: using jags

```
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
```

85

## 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
```

86

## 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 <- jags.model("file:///example1.jag", data=data, inits=inits, n.chains=2, n.adapt=500)  
params <- c("mu","sigmasq","y")  
samps <- coda.samples(jags,n, params, n.iter=2000)  
summary(samps)
```

87

## Example 1: using jags

Posterior predictive distribution

```
> summary(samps)

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

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

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

2. Quantiles for each variable:

          2.5%    25%    50%    75%   97.5%
mu     -0.182258 0.115228 0.242856 0.371231  0.647391
sigmasq  3.183236 3.764546 4.132446 4.580882  5.537943
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
```

88

## Analyzing FEV Data Set with Jags

```
<# Analysis of the FEV dataset using Jags
## read in data
data = read.table("https://raw.githubusercontent.com/chubb/SISC2017/master/data/fev.txt",
  col.names=c("seqn", "subjid", "age", "fev", "height", "sex", "smoke")
```

```
## now prepare data for Jags
data$age <- log(data$age)
data$smoke <- ifelse(data$smoke==0, age=mean(data$age),
  init=1, age=1, smoke=1)
init <- 1
model <- c("beta", "sigmasq", "ratiosq")
```

```
# define Jags model within R
model <- jags.model(file="fev.jag", data=data[,c], init=init, n.chains=2, n.adapt=500)
params <- c("beta", "sigmasq", "ratiosq")
fev.post <- coda.samples(model, params, n.iter=10000)
```

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

## convergence diagnostics
autocorr.plot(fev.post)
gelman.plot(fev.post)
geweke.diag(fev.post)
```

```
model{
  ## define likelihood of observations
  for (i in 1:N){
    y[i] ~ dnorm(mu[i], tauss)
    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)
  }
  tauss <- 1/sigmasq
  sigmasq ~ dunif(0,10)
  ## defining quantiles of interest (ratios of geometric means)
  for (i in 1:2){
    ratiosq[i] <- exp(beta[i+1])
  }
}
```

89

## 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]    0.82441 0.028043 1.939e-04   8.359e-04
beta[2]    0.08997 0.029832 2.109e-04   1.004e-03
beta[3]    0.09078 0.003063 2.166e-05   4.416e-05
ratiosq[1] 1.09463 0.032622 2.307e-04   1.096e-03
ratiosq[2] 1.09503 0.003354 2.371e-05   4.836e-05
sigmasq   0.04469 0.002473 1.749e-05   2.181e-05

2. Quantiles for each variable:

          2.5%    25%    50%    75%   97.5%
beta[1]    0.78038 0.81522 0.83430 0.85315 0.89011
beta[2]    0.03107 0.07004 0.09003 0.11047 0.14729
beta[3]    0.08472 0.08872 0.09077 0.09285 0.09678
ratiosq[1] 1.03156 1.07255 1.09420 1.11680 1.15869
ratiosq[2] 1.09841 1.09277 1.09502 1.09735 1.11012
sigmasq   0.04010 0.04297 0.04462 0.04632 0.04972
```

90



## Meta-analysis

### Efficacy of BCG Vaccine in the Prevention of Tuberculosis

#### Meta-analysis of the Published Literature

Simon A. Costello, MD; John J. Triccas, F. Steiner, MD; MPH; Catherine E. Seipio, OSC; Ma'hey E. Wilson, MD; Robert J. Rutledge, MD; Harriet V. Freling, MD; PhD; Federico Miesbod, 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 detail for evidence of vaccine allocation used to create control groups, equal numbers and location for treatment 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 tuberculosis deaths showed a protective effect from BCG vaccination of 71% (RR, 0.29; 95% CI, 0.16 to 0.52), and five studies reporting on meningitis showed a protective effect from BCG vaccine of 64% (OR, 0.38; 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 tuberculosis disease is greater than 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)



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

Source, y	Population		Cases of TB		TB Death			
	BCG	No BCG	BCG	No BCG	RR	BCG	No BCG	RR
Aronson, <sup>†</sup> 1946†	123	139	4	11	0.41	0	4	0.14
Ferguson and Simes, <sup>‡</sup> 1949	306	303	6	29	0.20	2	9	0.22
Rosenthal et al., <sup>§</sup> 1960§	231	220	3	11	0.26	0	4	0.12
Hart and Sutherland, <sup>¶</sup> 1977	13 598	12 867	62	248	0.24	...	...	...
Filippotti-Moller et al., <sup>**</sup> 1973	5059	5008	33	47	0.80	...	...	...
Stein and Aronson, <sup>**</sup> 1953	1541	1451	180	372	0.46	...	...	...
Vanderveen et al., <sup>**</sup> 1973	525	639	8	10	0.20	...	...	...
Medoff, <sup>†</sup> 1960§	86 391	68 391	505	499	1.01	...	...	...
Costeeze and Beerk, <sup>**</sup> 1966§	7499	7277	29	45	0.63	...	...	...
Rosenthal et al., <sup>**</sup> 1961§	1716	1666	17	65	0.25	1	6	0.16
Comstock et al., <sup>**</sup> 1974	50 634	27 338	186	141	0.71	8	12	0.36
Comstock and Webster, <sup>**</sup> 1969§	2486	2341	5	3	1.56	...	...	...
Comstock et al., <sup>**</sup> 1970§	16 913	17 854	27	29	0.98	...	...	...
Aronson et al., <sup>**</sup> 1960**	1541	1451	111	111	...	13	18	0.74
Levine and Sacks, <sup>**</sup> 1948††	568	528	...	...	...	6	6	0.93
Overall RR (95% confidence interval)			0.49 (0.34-0.70)			0.29 (0.18-0.53)		

\*RR indicates relative risk. Ellipses indicate data not reported.

†Infant study.

‡TB household.

§Data from a 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.)

\*\*Report on deaths.

††Follow-up sample sizes were not reported. We assumed follow-up was comparable in BCG and no BCG groups.

††Report on deaths is based on the same trial as Stein and Aronson, 1953.

††Data after 1952 recruitment.





## 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

94

---

---

---

---

---

---



## 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 assess 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.

95

---

---

---

---

---

---



## 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

96

---

---

---

---

---

---



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 estimate a common population effect.
    - Accounts for two sources of heterogeneity:
      - Within-study heterogeneity
      - Between-study heterogeneity

97



Systematic Reviews and Meta-Analysis

#### ■ Fixed-Effects (Mantel-Haenszel):

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

*k*: number of studies

$\kappa$ : number of studies

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

98



Systematic Reviews and Meta-Analysis

#### Random-Effects (DerSimonian-Laird):

$$Y_i = \mu_i + \sigma_i \varepsilon_i \text{ for } i=1,\dots,k$$

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

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

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

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

99

## 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!

100

---

---

---

---

---

---

---

---

---

## BCG Example

### Efficacy of BCG Vaccine in the Prevention of Tuberculosis

Meta-analysis of the Published Literature

Cochran A, Corrao MG, Doherty J, Ebrahim S, Gulliford S, Greenaway M, May E, Wilson M,  
Eliott E, Burridge M, Harvey V, Treworgy RD, Pocock S, Molesworth P, et al.

- 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 or without BCG vaccination
  - Heterogeneity among trials may be explained by geographic location and year
- Efficacy measure: Odds Ratio (OR)



101

---

---

---

---

---

---

---

---

---

## 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	Primdott-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	cpos	cneg
control group				

102

---

---

---

---

---

---

---

---

---

## 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(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))

```

103

## 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):

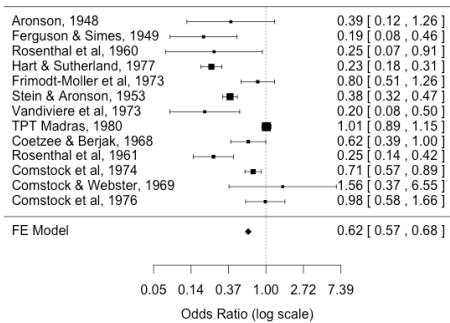
estimate    ci.lb    ci.ub
0.6229  0.5748  0.6750

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

```

104

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



105

## 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:

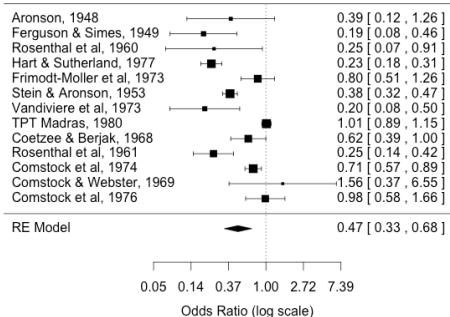
estimate      se     zval    pval ci.lb ci.ub
-0.7452  0.1860 -4.0057 <.0001 -1.1098 -0.3806 ***

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

The heterogeneity test shows strong evidence of heterogeneity in the 13 trials!<sup>106</sup>

## BCG Example (A): Standard Meta-Analysis

DerSimonian-Laird



107

## 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, transfexp)

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

108

## BCG Example: Meta-Regression Analysis

```

> res.m1

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%
R^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 (coefficients) (s):
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.

109

BCG Example (B): Bayesian Meta-Analysis

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

110

BCG Example (B): Bayesian Meta-Analysis

111

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

**Model 1 (B):**  $\delta_j \sim N(0, \nu_j^2), j=1, \dots, 13; \beta_i \sim N(0, \sigma_i^2)$

```
> fit1 = lola(Y ~ v1 + factor(trial) + factor(group), data=dat, Ntrials=N, family="binomial")
> summary(fit1)

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

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

Fixed effects:
mean   sd 0.025quant 0.5quant 0.975quant mode kld
factor(trial)1 -2.6017 0.2667 -23.5765 -2.5905 -2.1088 <-0.5674 0
factor(trial)2 -2.5823 0.1752 -22.3463 -2.5773 -2.2319 <-0.5672 0
factor(trial)3 -2.5823 0.1752 -22.3463 -2.5773 -2.2319 <-0.5672 0
factor(trial)4 -2.2176 0.0595 -43.3362 -4.2171 -4.1022 <-0.2159 0
factor(trial)5 -2.2176 0.0595 -43.3362 -4.2171 -4.1022 <-0.2159 0
factor(trial)6 -1.2584 0.0508 -13.3885 -1.2579 -1.1590 <-0.1574 0
factor(trial)7 -1.2584 0.0508 -13.3885 -1.2579 -1.1590 <-0.1574 0
factor(trial)8 -4.8537 0.0356 -45.0241 -4.8535 -4.3622 <-0.7729 0
4.8544 <-4.8532 -4.8537 0.0356 -45.0241 -4.8535 -4.3622 <-0.7729 0
factor(trial)9 -5.0772 0.1177 -55.1514 -5.0747 -4.8529 <-0.0996 0
factor(trial)10 -0.4792 0.1131 -37.7075 -0.4770 -0.2620 <-0.4725 0
factor(trial)11 -0.4792 0.1131 -37.7075 -0.4770 -0.2620 <-0.4725 0
factor(trial)12 -6.1843 0.3541 -66.3974 -6.1635 -5.5465 <-0.1201 0
factor(trial)13 -6.1843 0.3541 -66.3974 -6.1635 -5.5465 <-0.1201 0
factor(group)1 -0.4784 0.0413 -37.5987 -0.4784 -0.3973 <-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: -236.43

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).
```

112

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

**Model 2 (B):**  $\delta_j \sim N(0, \nu_j^2), j=1, \dots, 13;$   
 $\nu^2 \sim \text{IG}(a, b); \beta_i \sim N(0, \sigma_i^2)$

```
> fit2 = lola(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 time Post-processing      Total
0.0983           0.0381           0.0551           0.1825

Fixed effects:
mean   sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) -4.2043 0.4260 -5.0507 -4.2041 -3.3600 <-0.2339 0
factor(group)1 -0.4785 0.0413 -0.5598 -0.4785 -0.3976 <-0.4784 0

Random effects:
None. This is a 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
Precision for group 0.4633 0.1930 0.1802 0.4335 0.9236 0.3733

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

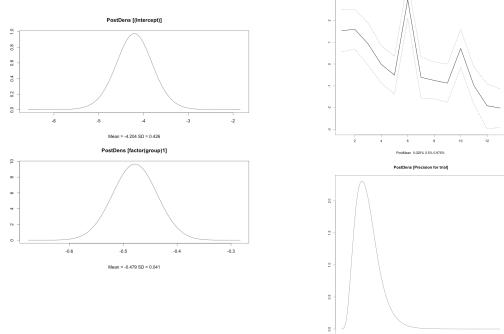
Marginal Likelihood: -209.55

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

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

113

**Model 2 (B):**



114

**Model 3 (B):**

```

logit( $p_i$ ) =  $\delta_0 + \beta_1 Group + \beta_2 Latitude + \beta_3 Group * Latitude,$ 
 $\delta_j \sim N(0, \sigma^2), j = 1, \dots, 13; \sigma^2 \sim IG(a, b);$ 
 $\beta_k \sim N(0, \sigma_k^2); k = 1, 2, 3$ 

> summary(fit3)

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

Fixed effects:
            mean     sd 0.025quant 0.5quant 0.975quant mode kid
(Intercept) -14.891 0.3571 -4.8511 0.1398 -3.4323 4.1398 0
factor(group1) -0.7166 0.0480 -0.7164 -0.7164 -0.7164 0
centeredLatitude 0.0736 0.0256 0.0227 0.0736 0.1246 0.0736 0
factor(group1):centeredLatitude -0.0334 0.0028 -0.0388 -0.0333 -0.0279 -0.0333 0

Random effects:
Name    Model
trial   IID model

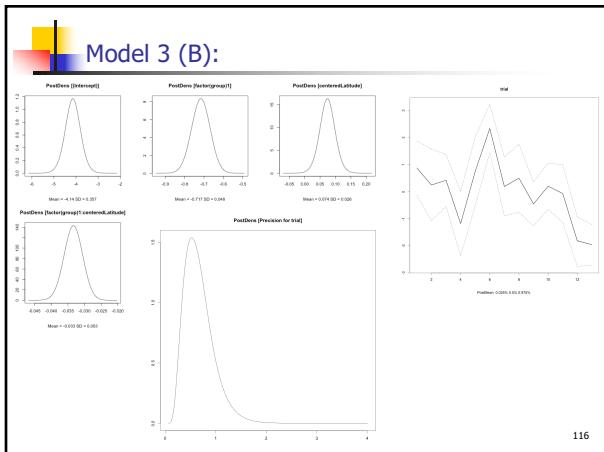
Model hyperparameters:
mean     sd 0.025quant 0.5quant 0.975quant mode
Precision for trial 0.6693 0.2933 0.2461 0.6220 1.3742 0.5249

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.

115



116

**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

117

**Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons**

Hsunhee Hong, MS, Bradley P. Gordin, PhD, Tolvana A. Shanliyan, MD, MS,  
Jean F. Wyman, PhD, Reina Ramakrishnan, MPH, François Saïnfort, 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 outcome data. **Design.** We searched major electronic bibliographic databases, FDA 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 placebo as better than all products. **Conclusion.** Bayesian methods are 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. **Keywords:** nephrology; Bayesian meta-analysis; comparative effectiveness; systematic reviews; hierarchical models. (*Med Decis Making* 2013;33:702–714)

118

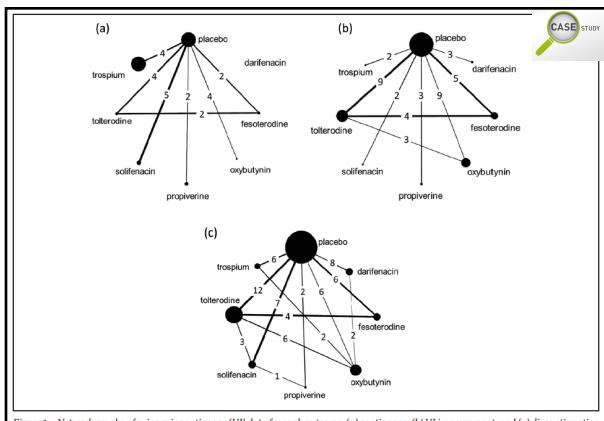


Figure 1 Network graphs of urinary incontinence (UI) data for each outcome: (a) continence, (b) UI improvement, and (c) discontinuation due to adverse events (AEs). The size of each node represents the number of studies investigating the drug, and the thickness of each edge implies the total number of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

120



## 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.

121

---

---

---

---

---

---



## 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:

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

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

122

---

---

---

---

---

---



## 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

123

---

---

---

---

---

---

## 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	9	140	1
2	1	3	23	140
3	1	4	10	138
4	2	2	11	78
5	2	3	12	85
6	2	4	29	170
7	3	1	75	31
8	3	3	36	714
9	4	1	2	106
10	4	3	9	205
11	5	1	58	549
12	5	3	237	1561
13	6	1	0	33
14	6	3	9	48
15	7	1	1	100
16	7	3	31	98
17	8	1	1	31
18	8	3	26	95
19	9	1	6	39
20	9	3	17	77

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

124

## 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

125

## Meta-Analysis: Mixed and Indirect Treatment Comparisons

- Fixed Effects
- Random Effects

$r_{jk} \sim \text{Binomial}(p_{jk}, N_{jk})$

$\text{logit}(p_{jk}) = \mu_j + \delta_{XY} I_{(k=Y)}$

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

$\mu_j, d_{AB}, d_{AC}, d_{AD} \sim N(0, 100^2)$

$r_{jk} \sim \text{Binomial}(p_{jk}, N_{jk})$

$\text{logit}(p_{jk}) = \mu_j + \delta_{XY} I_{(k=Y)}$

$\delta_{XY} \sim N(d_{XY}, \sigma^2)$

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

$\mu_j, d_{AB}, d_{AC}, d_{AD} \sim N(0, 100^2)$   
 $\sigma \sim U(0, 2)$

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)

126

## 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	$d[2] - d[1] = d[2]$
1,3	3	1	$d[3] - d[1] = d[3]$
1,4	4	1	$d[4] - d[1] = d[4]$
2,3	3	2	$d[3] - d[2]$
2,4	4	2	$d[4] - d[2]$
3,4	4	3	$d[4] - d[3]$

127

## 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:
        - $\text{Logit}(T_k) = A + d_k$

128

## Meta-Analysis: Mixed and Indirect Treatment Comparisons

```
## Fixed Effects model
model {
  # loop over 50 observations
  for(i in 1:50) {
    # i=1 is baseline comparison
    # Response[i] ~ dbin(p[1],n[i])
    Response[i] ~ dnorm(mu[1],tau[1])
    delta[i] <- mu[1] - mu[2*(i-1)+1]
    delta[i] <- d[Treatment[i]] - d[Baseline[i]]
  }
  # vague priors for intercepts (effect for baseline comparison group)
  for (i in 1:24) {mu[i] ~ dnorm(0,.0001)}
  # assume Treatment 1 as 0 reflects of other Treatments is relative to this Treatment 1
  d[1] <- 0
  # vague priors for 3 basic treatment effect parameters
  for (i in 2:3) {d[i] ~ dnorm(0,.001)}
}

# Absolute treatment effects
# prior precision for Treatment 1, adv=.38
prior d[1] ~ dnorm(0,.38)
# external info on A,
A ~ dnorm(-2.6,predA)

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

#rank the treatment effects (with 1-beat) & record the best treatment
ix <- rank(d[1:4])
beat <- max(ix[1:4])

#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:c-1) {
    logr[c,k] <- d[k] - d[c]
  }
}
```

129

## Meta-Analysis: Mixed and Indirect Treatment Comparisons

```

# random effects model
for(i in 1:nrow(datajag)) {
  # loop over 50 observations
  for(j in 1:50) {
    # likelihood
    logit(p[i,j]) <- mu[Study[i],] + delta[i]*(I-equals[Treatment[i], Baseline[i]])
    # draw effect from random effects distribution
    delta[i] <- rnorm(1,0,0.01)
    # population mean effect
    mu[i] <- d1[Treatment[i]] - d1[baseline[i]]
    # vague prior for intercept (baseline group)
    delta[i] <- rnorm(1,0,0.001)
    mu[i] ~ dnorm(0, 2000)
    # net effect of treatment i as 0 (all other treatment effects are relative to this one)
    d1[Treatment[i]] <- mu[i]
    # flat priors for 3 basic treatment parameters
    d1[baseline] ~ dnorm(0, 0001)
    d1[delta] ~ dnorm(0, 0001)

    # Absolute treatment effects
    # Prior for intercept (baseline treatment 1, sd=38
    precA <- pow(38,-2)
    A = dnorm(0, precA)
    d1[baseline] ~ dnorm(A, d1[baseline])
    logit(p[i,j]) <- A + d1[i]

    ## prior for variance component
    # prior for sigma^2
    sd ~ dmf(d1[2])
    # Prior for treatment effects (with 1-beat) & record the best treatment
    rk <- 5 - rank(d1[2])
    best[1] <- d1[2]
    # All pairwise log odds ratios and odds ratios (some of these calculations are redundant, but needed to run)
    for(k in 1:i-1) {
      for(l in k+1:i) {
        logit(rk[k,l]) <- A + d1[k]
      }
    }
  }
}

## 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="mtb-fe.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
parameters <- c("d", "lor", "rk", "best", "T")
post1 <- coda.samples(model1, parameters, 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="mtb-re.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
parameters <- c("d", "lor", "rk", "best", "sd")
post2 <- coda.samples(model2, parameters, n.iter=10000)
summary(post2)

```

130

## 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="mtb-fe.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
parameters <- c("d", "lor", "rk", "best", "T")
post1 <- coda.samples(model1, parameters, 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="mtb-re.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
parameters <- c("d", "lor", "rk", "best", "sd")
post2 <- coda.samples(model2, parameters, n.iter=10000)
summary(post2)

```

131

## Meta-Analysis: Mixed and Indirect Treatment Comparisons

```

> summary(post1)
Iterations = 501:15000
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     Std. Error      Std. Dev.
Mean    0.026250  0.0002400  0.002400
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
best[1] 0.000000  0.000000  0.000000
best[2] 0.000000  0.000000  0.000000
best[3] 0.000000  0.000000  0.000000
best[4] 0.000000  0.000000  0.000000
d[1]   0.2729  0.1283  0.003923
d[2]   0.2729  0.1283  0.003923
d[3]   0.2729  0.1283  0.003923
d[4]   0.2729  0.1283  0.003923
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
lor[1,1] 0.049254  0.033483  0.002988
lor[2,1] 0.049254  0.033483  0.002988
lor[3,1] 0.049254  0.033483  0.002988
lor[4,1] 0.049254  0.033483  0.002988
lor[1,2] 0.000000  0.000000  0.000000
lor[2,2] 0.000000  0.000000  0.000000
lor[3,2] 0.000000  0.000000  0.000000
lor[4,2] 0.000000  0.000000  0.000000
lor[1,3] 0.000000  0.000000  0.000000
lor[2,3] 0.000000  0.000000  0.000000
lor[3,3] 0.000000  0.000000  0.000000
lor[4,3] 0.000000  0.000000  0.000000
lor[1,4] 0.000000  0.000000  0.000000
lor[2,4] 0.000000  0.000000  0.000000
lor[3,4] 0.000000  0.000000  0.000000
lor[4,4] 0.000000  0.000000  0.000000
rk[1,1] 0.049254  0.033483  0.002988
rk[2,1] 0.049254  0.033483  0.002988
rk[3,1] 0.049254  0.033483  0.002988
rk[4,1] 0.049254  0.033483  0.002988
rk[1,2] 0.000000  0.000000  0.000000
rk[2,2] 0.000000  0.000000  0.000000
rk[3,2] 0.000000  0.000000  0.000000
rk[4,2] 0.000000  0.000000  0.000000
rk[1,3] 0.000000  0.000000  0.000000
rk[2,3] 0.000000  0.000000  0.000000
rk[3,3] 0.000000  0.000000  0.000000
rk[4,3] 0.000000  0.000000  0.000000
rk[1,4] 0.000000  0.000000  0.000000
rk[2,4] 0.000000  0.000000  0.000000
rk[3,4] 0.000000  0.000000  0.000000
rk[4,4] 0.000000  0.000000  0.000000
best[1] 0.000000  0.000000  0.000000
best[2] 0.000000  0.000000  0.000000
best[3] 0.000000  0.000000  0.000000
best[4] 0.000000  0.000000  0.000000
d[1]   0.2729  0.1283  0.003923
d[2]   0.2729  0.1283  0.003923
d[3]   0.2729  0.1283  0.003923
d[4]   0.2729  0.1283  0.003923
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283  0.003923
l1[2,1] -0.02790  0.1283  0.003923
l1[3,1] -0.02790  0.1283  0.003923
l1[4,1] -0.02790  0.1283  0.003923
l1[1,2]  0.000000  0.000000  0.000000
l1[2,2]  0.000000  0.000000  0.000000
l1[3,2]  0.000000  0.000000  0.000000
l1[4,2]  0.000000  0.000000  0.000000
l1[1,3]  0.000000  0.000000  0.000000
l1[2,3]  0.000000  0.000000  0.000000
l1[3,3]  0.000000  0.000000  0.000000
l1[4,3]  0.000000  0.000000  0.000000
l1[1,4]  0.000000  0.000000  0.000000
l1[2,4]  0.000000  0.000000  0.000000
l1[3,4]  0.000000  0.000000  0.000000
l1[4,4]  0.000000  0.000000  0.000000
l1[1]  0.049254  0.033483  0.002988
l1[2]  0.049254  0.033483  0.002988
l1[3]  0.049254  0.033483  0.002988
l1[4]  0.049254  0.033483  0.002988
l1[1,1] -0.02790  0.1283 
```

## Meta-Analysis: Mixed and Indirect Treatment Comparisons

133



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

## Diagnostic testing with missing gold standard

134

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**

---

Serology	Stool examination					
	+	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
Stool examination	0.50	0.03–0.98	0.74	0.41–0.98	0.80	0.23–0.99	0.76	0.52–0.91
S <sub>1</sub>	0.24	0.07–0.47	0.30	0.21–0.47	—	—	0.31	0.22–0.44
C <sub>1</sub>	0.95	0.89–0.99	0.95	0.88–0.99	—	—	0.96	0.91–0.99
PPV <sub>1</sub>	0.84	0.10–1.00	0.95	0.74–1.00	—	—	0.98	0.88–1.00
NPV <sub>1</sub>	0.56	0.03–0.98	0.33	0.02–0.73	—	—	0.30	0.11–0.63
Serology	S <sub>2</sub>	0.81	0.63–0.92	—	0.83	0.73–0.92	0.89	0.80–0.95
	C <sub>2</sub>	0.72	0.31–0.98	—	0.58	0.22–0.94	0.67	0.36–0.95
	PPV <sub>2</sub>	0.76	0.07–1.00	—	0.91	0.18–1.00	0.90	0.82–1.00
	NPV <sub>2</sub>	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  $2 \times 2$  table**

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

$Y_1$  and  $Y_2$  are latent/unobserved data

136

## Reproducing analyses: Using only one diagnostic test

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

$$PPT = P(T+) = P(T+|D)P(D) + P(T+|D^c)P(D^c)$$

$$= S\pi + (1 - C)(1 - \pi)$$

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

137

## 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-D)*(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.

138

## Reproducing analyses: Using only one diagnostic test

```

data = list(N=162, a=40, Y1=NA, Y2=NA, ap=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)])

```

Posterior Estimates

	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	24.4454211	35.9750000	85.0000000	120.0000000
pi	0.7401364	0.16335664	0.4075330	0.7581660	0.9855299

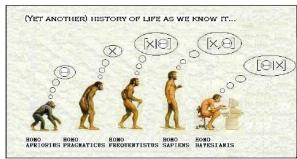
139

## Final Comments

There is 'art' in Bayesian Analysis



- Achieving 'mastery' requires practice!



140