Lecture 9.
Hierarchical Regression Models

Learning Outcomes

After this session students should be able to:

- Explain how a hierarchical structure can be applied to regression models;
- Describe how hierarchical models can be extended to include more than two levels and cross-classified data;
- Use hierarchical models for modelling variances.

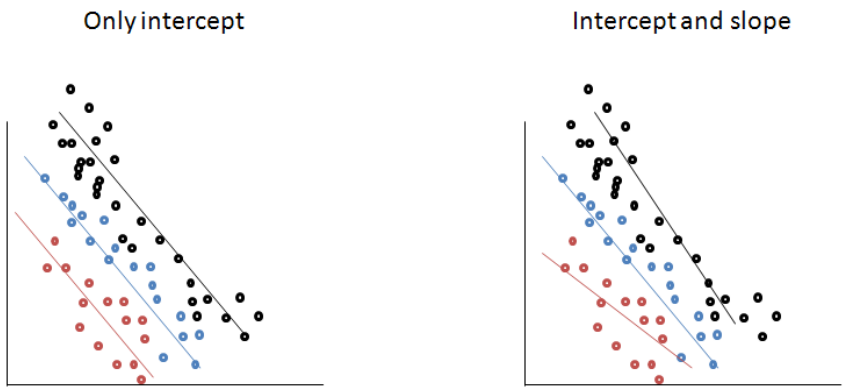
Hierarchical regression models

- Hierarchical structure can be applied to regression models
- One intercept and/or slope for each group (e.g. individual) which are then exchangeable

e.g. linear regression

- $i = 1, \ldots, I_j$ groups, $j = 1, \ldots, J$ measurements
  $\rightsquigarrow y_{ij} \sim N(\mu_{ij}, \sigma^2) \qquad \mu_{ij} = \alpha_i + \beta_i x_{ij}$

$$y_{ij} \sim N(\alpha_i + \beta x_{ij}, \sigma^2)$$
$$\alpha_i \sim N(m_\alpha, s_\alpha^2)$$
$$\beta \sim N(0, 10000)$$



$$y_{ij} \sim N(\alpha_i + \beta_i x_{ij}, \sigma^2)$$
$$\alpha_i \sim N(m_\alpha, s_\alpha^2)$$
$$\beta_i \sim N(m_\beta, s_\beta^2)$$

Example: Hepatitis B Immunisation

Background

- Hepatitis B (HB) is endemic in Africa
- National program of childhood vaccination against HB introduced in Gambia
- Program effectiveness depends on duration of immunity afforded by vaccination

Data

- 106 children immunized against HB
- For each child: anti-HB titre measured at time of vaccination (baseline) and on 2 or 3 follow-up occasions

Study objective

- To obtain a model useful for predicting an individual child's protection against HB after vaccination

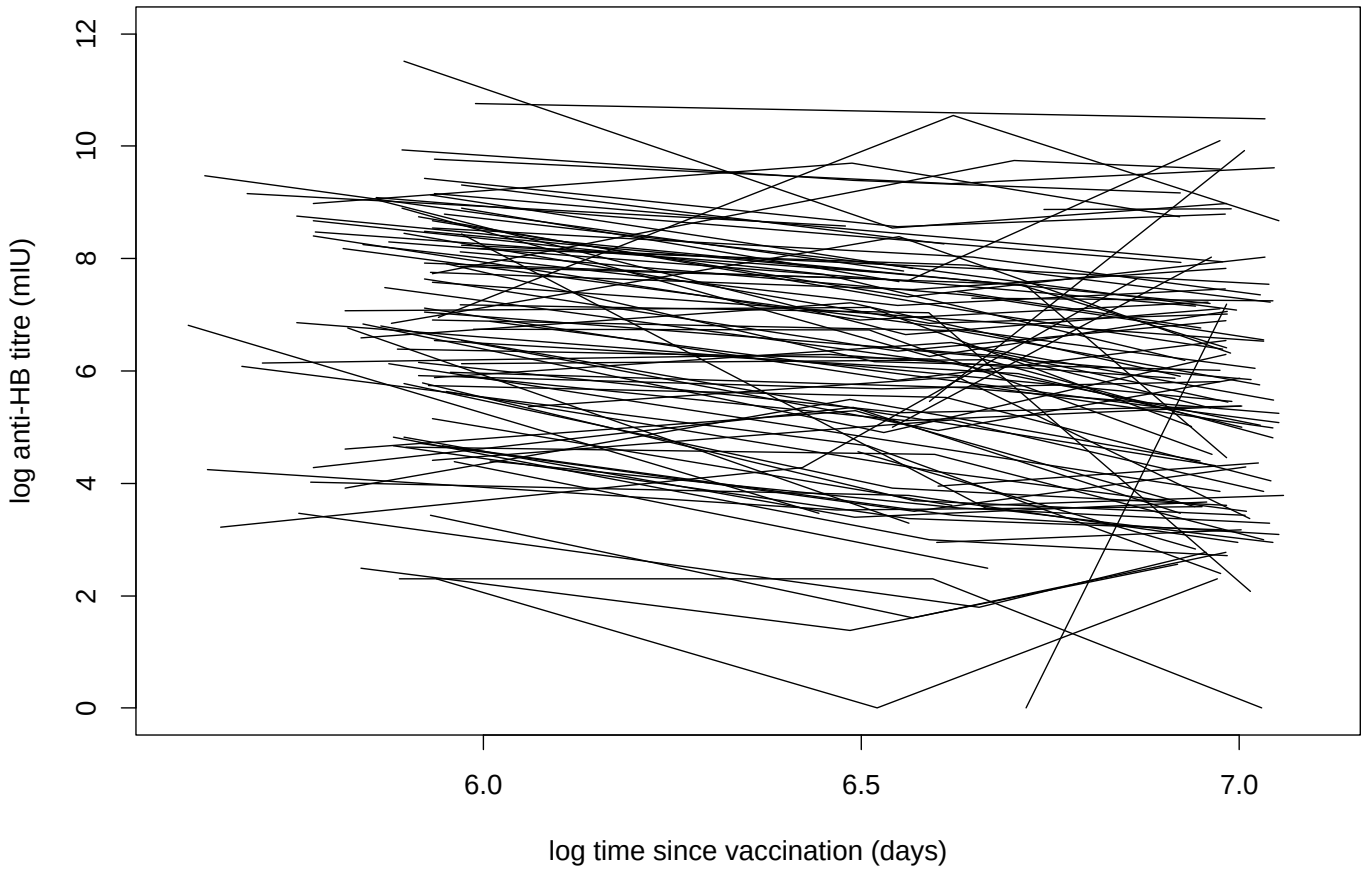
Related studies

- Similar study in Senegal found:

$$\text{anti-HB titre} \propto \frac{1}{T}$$

where T = time since HB vaccination

Raw data for a subset of 106 individuals



Non hierarchical linear model (LM) for the HB data

1. Probability distribution(likelihood) for responses:

$$y_{ij} \sim \text{Normal}(\mu_{ij}, \sigma^2)$$

where y_{ij} = log of the jth anti-HB titre measurement for child i

2. Linear predictor:

$$\mu_{ij} = \alpha + \beta(t_{ij} - \bar{t}) + \gamma(y_{0i} - \overline{y_0})$$

where

t_{ij} = log of time (days since vaccination) of jth measurement for child i

y_{0i} = log of baseline anti-HB titre for child i

Problems

- Assumes a common regression line for all children
- Takes no account of the repeated measurements within children

$\Rightarrow$ modify LM to allow separate intercept and slope for each child:

$$\begin{aligned} y_{ij} &\sim \text{Normal}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \alpha_i + \beta_i(t_{ij} - \bar{t}) + \gamma(y_{0i} - \overline{y_0}) \end{aligned}$$

Assumes that *conditionally* on α_i and β_i, $\{y_{ij}, j = 1, 2, ..\}$ are independent

- Assume α_i's are exchangeable and β_i's are exchangeable, e.g.

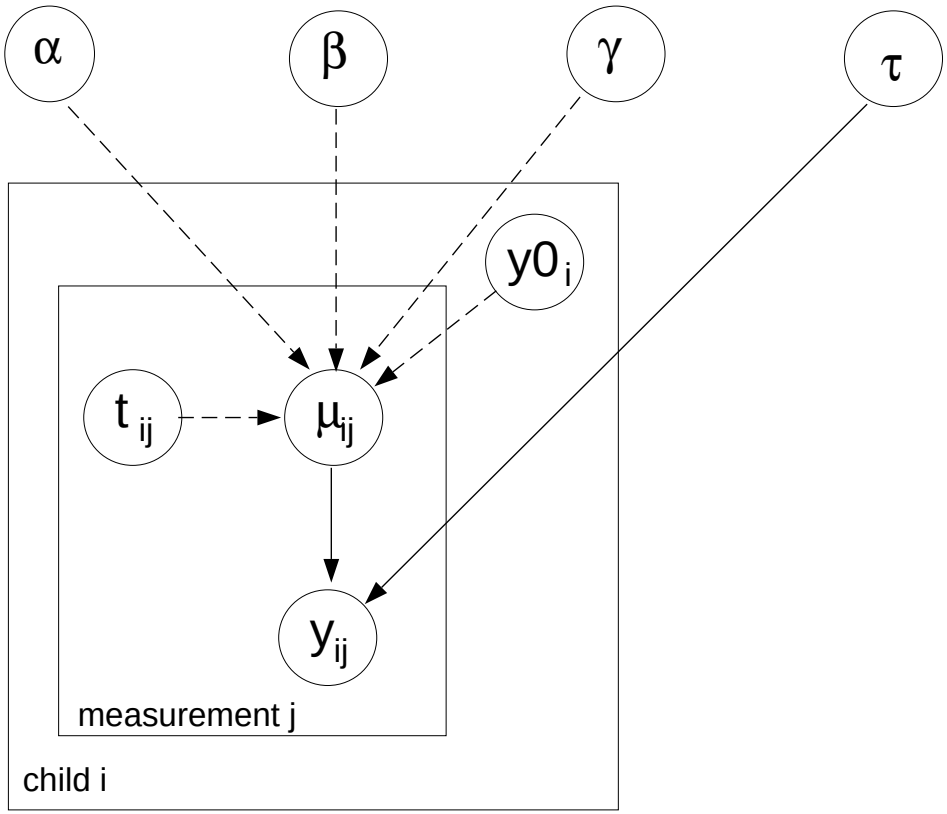
$$\begin{aligned} \alpha_i &\sim \text{Normal}(\mu_\alpha, \sigma^2_\alpha) \quad i = 1, ..., 106 \\ \beta_i &\sim \text{Normal}(\mu_\beta, \sigma^2_\beta) \quad i = 1, ..., 106 \end{aligned}$$

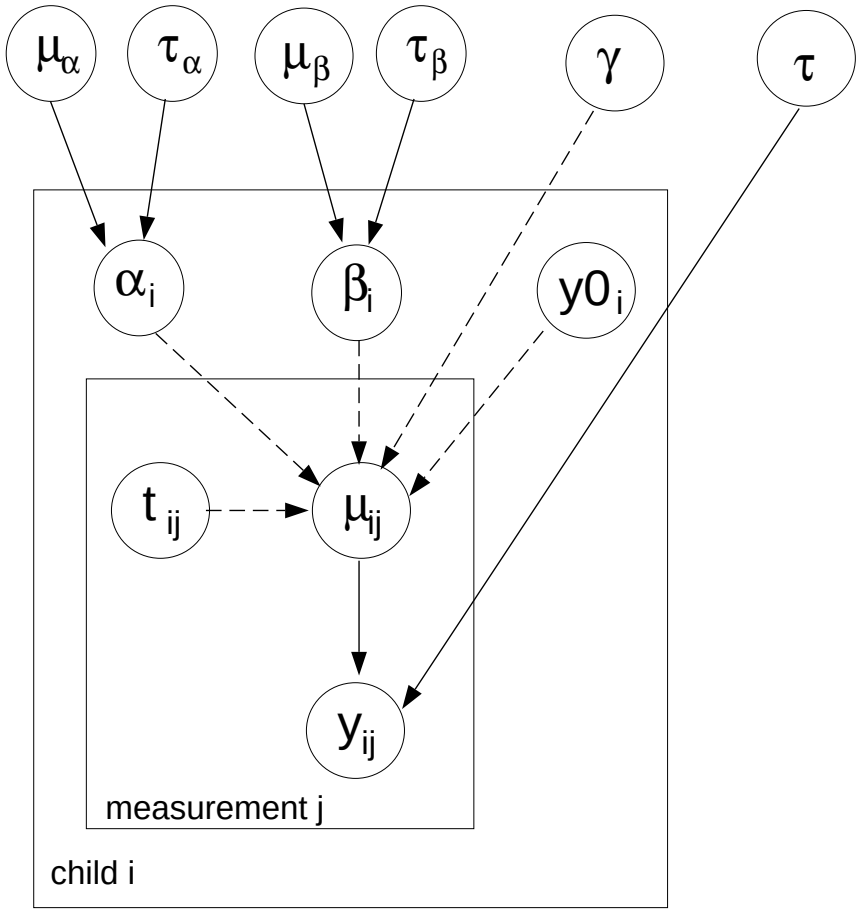
- We may then assume vague priors for the *hyperparameters* of the population distribution, e.g.

$$\begin{aligned} \mu_\beta, \mu_\alpha &\sim \text{Normal}(0, 10000) \\ \tau_\alpha = \sigma_\alpha^{-2}, \tau_\beta = \sigma_\beta^{-2} &\sim \text{Gamma}(0.001, 0.001) \end{aligned}$$

This is an example of a *Hierarchical LM* or *Linear Mixed Model (LMM)* or *Random Coefficients* model

Graph of a LM and LMM for the HB data





Implementation in WinBUGS

Data contain 2 or 3 observations per child, so ragged array again

Child	Log Titre	Log Time
1	4.99, 8.02	6.54, 6.96
2	6.83, 4.91, 6.29	5.84, 6.52, 6.98
3	3.95, 4.35	6.60, 7.02
4	....	.....

Model code using nested index formulation

```
for(k in 1:TotalObs) {
  y[k] ~ dnorm(mu[k],tau)
  mu[k] <- alpha[child[k]]+ beta[child[k]]*(t[k]-tbar)
        + gamma*(y0[child[k]]-y0bar)
}
for(i in 1:N) {
  alpha[i] ~ dnorm(mu.alpha, tau.alpha)
  beta[i] ~ dnorm(mu.beta, tau.beta)
}
.... etc.
```

Data

```
list(N=106, TotalObs=288)
```

y[]	t[]	child[]	y0[]
4.99	6.54	1	8.61
8.02	6.96	1	8.61
6.83	5.84	2	7.10
4.91	6.52	2	7.10

```
.....
END
```

Results for the LM and LMM models fitted to the HB data

Parameters	LM	Parameters	LMM
α	6.03 (0.10)	μ_α	6.04 (0.15)
β	-1.05 (0.22)	μ_β	-1.08 (0.13)
γ	0.67 (0.06)	γ	0.67 (0.08)
σ^2	3.00 (0.26)	σ^2	1.01 (0.11)
		σ^2_α	2.02 (0.35)
		σ^2_β	0.06 (0.09)
DIC	1136	DIC	913
p_D	4.0	p_D	95.1

Note how the residual variance σ^2 has been reduced, being explained partially by α and β

Multiple random effects and cross classified data

- Straightforward to extend basic 2-level hierarchical model to include multiple random effects at different levels:
  - nested hierarchies, e.g. THM measurements within zones within regions; pupils within classes within schools
  - cross-classified hierarchies, e.g. THM measurements cross-classified within zones and years; pupils cross-classified within primary and secondary schools
- Easiest to formulate cross-classified models in BUGS using nested index notation

Example: Schools – exam scores cross-classified by primary and secondary school

- We use a random sample of 800 children who attended 132 primary schools and 19 secondary schools in Scotland
- The following variables were used

`Y`:	Exam attainment score of pupils at age 16
`VRQ`:	verbal reasoning score taken on secondary school entry
`SEX`:	Pupil's gender (0 = boy, 1 = girl)
`PID`:	Primary school identifying code
`SID`:	Secondary school identifying code

- A normal hierarchical model is fitted, with independent random effects for primary school and secondary school
- Verbal reasoning score and gender are included as 'fixed' covariate effects (but note that in Bayesian framework, 'fixed' effect coefficients are still assigned prior distributions)

Model:

$$
\begin{aligned}
i &= 1, \ldots, N \\
y_i &\sim \text{Normal}(\mu_i, \sigma^2_{[y]}) \\
\mu_i &= \alpha + \beta_1 \text{SEX}_i + \beta_2 \text{VRQ}_i + \theta_{[ps]\text{PID}_i} + \theta_{[ss]\text{SID}_i} \\
\alpha &\sim N(0, 10000) \\
\beta_1 &\sim N(0, 10000) \\
\beta_2 &\sim N(0, 10000) \\
j &= 1, \ldots, J \text{ Primary school} \\
\theta_j &\sim N(0, \sigma^2_{ps}) \\
k &= 1, \ldots, K \text{ Secondary school} \\
\theta_k &\sim N(0, \sigma^2_{ss})
\end{aligned}
$$

BUGS model code

```
for(i in 1:Nobs) {
 y[i] ~ dnorm(mu[i], tau.y)
 mu[i] <- alpha + beta[1]*SEX[i] + beta[2]*VRQ[i] +
                  theta.ps[PID[i]] + theta.ss[SID[i]]
}
# random effects distributions (note: non centered)

# primary school effects
for (j in 1:Nprim) { theta.ps[j] ~ dnorm(0, tau.ps) }

# secondary school effects
for (k in 1:Nsec) { theta.ss[k] ~ dnorm(0, tau.ss) }

# regression coefficients
alpha ~ dnorm(0, 0.000001)    # intercept
for(q in 1:2) { beta[q] ~ dnorm(0, 0.000001)   }
```

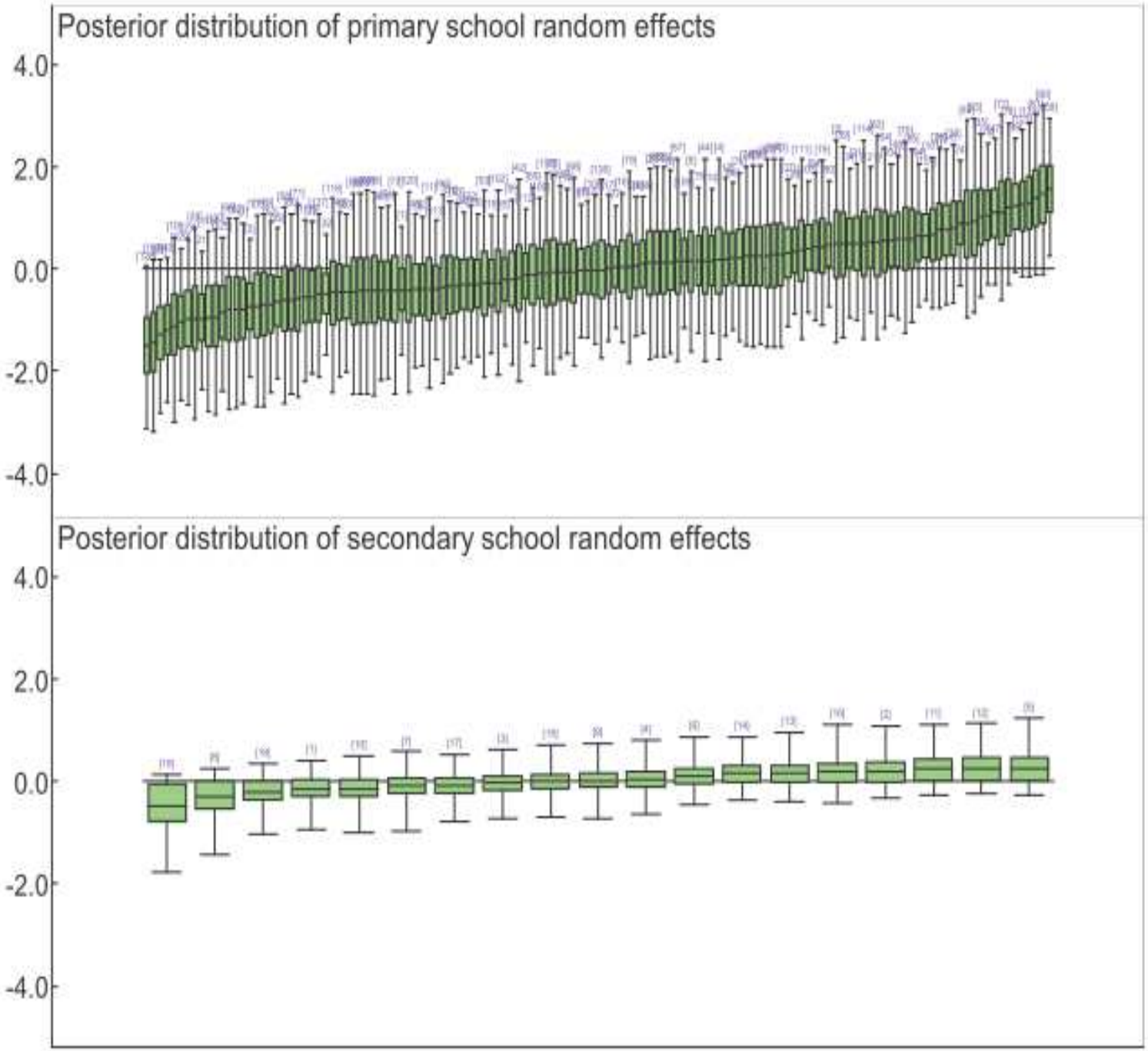
Variances

```
   priors on regression coefficients and variances
tau.y ~ dgamma(0.001, 0.001)
sigma2.y <- 1/tau.y              # residual error variance
tau.ps ~ dgamma(0.001, 0.001)
sigma2.ps <- 1/tau.ps            # between primary school variance
tau.ss ~ dgamma(0.001, 0.001)
sigma2.ss <- 1/tau.ss            # between secondary school variance

# percentage of total variance explained ...
#..by primary school effects
VPC.ps <- sigma2.ps / (sigma2.y + sigma2.ps + sigma2.ss)
#..by secondary school effects
VPC.ss <- sigma2.ss / (sigma2.y + sigma2.ps + sigma2.ss)
```

Results

Parameters	Model 1		Model 2	
α	5.53	(5.17, 5.88)	5.85	(5.59, 6.10)
β_1 (sex)	–	–	0.23	(-0.08, 0.53)
β_2 (VRQ)	–	–	0.16	(0.15, 0.17)
$\sigma^2_{[y]}$	8.18	(7.35, 9.10)	4.49	(4.03, 5.00)
$\sigma^2_{[ps]}$	1.12	(0.43, 1.98)	0.36	(0.08, 0.70)
$\sigma^2_{[ss]}$	0.19	(0.10, 0.82)	0.02	(0.0007, 0.12)
VPC_{ps}	11.8%	(4.7%, 19.8%)	7.4%	(1.5%, 13.8%)
VPC_{ss}	2.0%	(0.1%, 8.3%)	0.4%	(0.01%, 2.4%)
DIC	4008		3514	
p_D	58.0		43.8	


Posterior distribution of primary school random effects
4.0
2.0
0.0
-2.0
-4.0
Posterior distribution of secondary school random effects
4.0
2.0
0.0
-2.0
-4.0

Heteroscedasticity

- Heteroscedasticity $\rightarrow$ non constant variance
- Can occur at any level of hierarchical model
- Easily handled in MCMC framework by modelling variance as a specified function of other variables

Example: complex level 1 variation in Schools example

Original model:

$$\begin{aligned} y_i &\sim \text{Normal}(\mu_i, \sigma^2_{[y]}) \\ \mu_i &= \alpha + \beta_1 \text{SEX}_i + \beta_2 \text{VRQ}_i + \theta_{[ps]\text{PID}_i} + \theta_{[ss]\text{SID}_i} \\ &.... \end{aligned}$$

Complex level 1 variation depending on VRQ:

$$\begin{aligned} y_i &\sim \text{Normal}(\mu_i, \sigma^2_{[y]i}) \\ \log \sigma^2_{[y]i} &= \gamma_1 + \gamma_2 \text{VRQ}_i \\ \mu_i &= ..... \end{aligned}$$

Along with priors on α, β_1, β_2 and random effects variances, also need priors on coefficients of variance model:

$$\gamma_1 \sim \text{Normal}(0, 10000)$$

$$\gamma_2 \sim \text{Normal}(0, 10000)$$

BUGS model code

```
for(i in 1:Nobs) {
 Y[i] ~ dnorm(mu[i], tau.y[i])
 mu[i] <- alpha + beta[1]*SEX[i] + beta[2]*VRQ[i] +
                  theta.ps[PID[i]] + theta.ss[SID[i]]

 # complex level 1 variance
 logsigma2.y[i] <- gamma[1] + gamma[2]*VRQ[i]
 tau.y[i] <- 1/exp(logsigma2.y[i])
}

# remaining code is same as before
......
......
# except no longer need prior on residual error variance
##tau.y ~ dgamma(0.001, 0.001)
##sigma2.y <- 1/tau.y              # residual error variance

# instead need to include priors on coefficient of variance model
for(k in 1:2) { gamma[k] ~ dnorm(0, 0.000001)   }
```

BUGS model code continued....

```
## VPC will now depend on value of VRQ

# level 1 variance for child with VRQ in lowest 10th percentile
sigma2.y.lowVRQ <- exp(gamma[1] + gamma[2] * (-19))

# level 1 variance for child with VRQ in highest 10th percentile
sigma2.y.hiVRQ <- exp(gamma[1] + gamma[2] * 15)


## percentage of total variance explained by primary school effects....

# .....for pupils with low VRQ
VPC.ps.lowVRQ <- sigma2.ps / (sigma2.y.lowVRQ + sigma2.ps + sigma2.ss)

# .....for pupils with hi VRQ
VPC.ps.hiVRQ <- sigma2.ps / (sigma2.y.hiVRQ + sigma2.ps + sigma2.ss)
```

Initial values

- Remember to edit initial values from previous model to:
  - remove initial values for `tau.y`
  - add initial values for `gamma` vector
- Some care needed when specifying initial values for `gamma[2]` to avoid numerical problems in BUGS
  - `gamma[2]` measures effect of unit change in VRQ (which ranges from -30 to 40) on log residual variance
  - Residual variance was around 5 from previous analysis, so expect values of log variance around $\log 5 = 1.6$

  $\Rightarrow$ `gamma[2]` should be quite small ($<< 1$)

e.g.

```
list(alpha = 0, tau.ps = 1, tau.ss = 1, beta=c(0,0),
gamma=c(1, 0.001))
```

Results

Parameter	Posterior mean	95% CI
γ_2	0.019	(0.008, 0.029)
VPC_{ps} (low VRQ)	9.0%	(2.0%, 18.0%)
VPC_{ps} (hi VRQ)	5.0%	(1.0%, 10.4%)
VPC_{ss} (low VRQ)	0.6%	(0.01%, 3.3%)
VPC_{ss} (hi VRQ)	0.4%	(0.01%, 1.9%)
DIC	3503	
p_D	43.3	

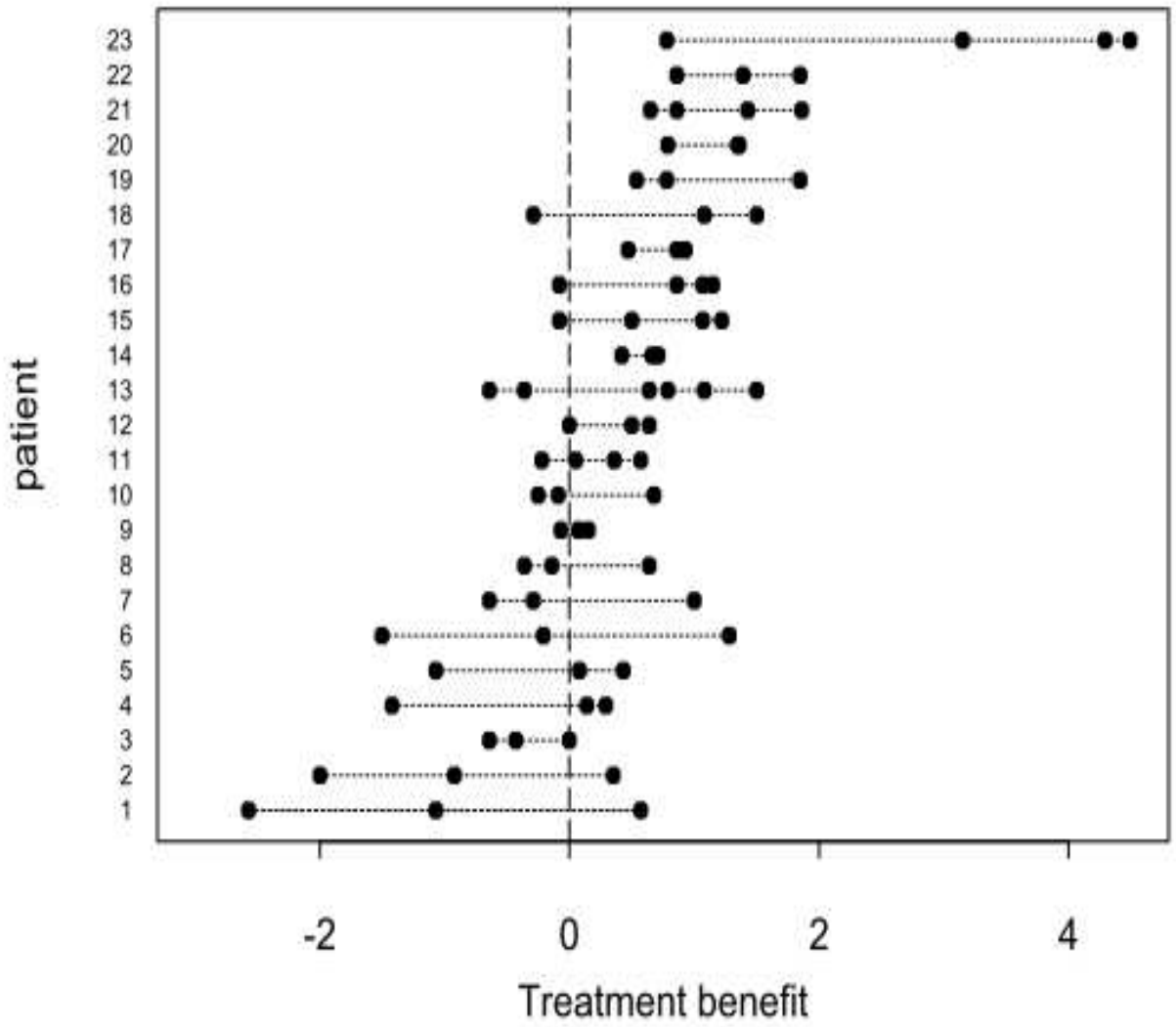
Recall model with homoscedastic level 1 variance had DIC = 3514, p_D = 43.8, so heteroscedastic model preferred

Hierarchical models for variances – Example: N-of-1 trials

Spiegelhalter et al (2004) Example 6.10

- N-of-1 trials $\rightarrow$ repeated within-person crossover trials
- Often suitable for investigating short-term symptom relief in chronic conditions
- Example:
  - **Intervention:** Amitriptyline for treatment of fibromyalgia to be compared with placebo.
  - **Study design:** 23 N-of-1 studies - each patient treated for a number of periods (3 to 6 per patient), and in each period both amitriptyline and placebo were administered in random order
  - **Outcome measure:** Difference in response to a symptom questionnaire in each paired crossover period. A positive difference indicates Amitriptyline is superior
  - **Evidence from study:** 7/23 experienced benefit from the new treatments in all their periods

Raw data for each patient



Statistical model

If y_{kj} is the j^{th} measurement on the k^{th} individual, we assume $y_{kj} \sim \mathrm{N}(\theta_k, \sigma_k^2)$

Assume both θ_k's and σ_k^2's are *exchangeable*, in the sense there is no reason to expect systematic differences and we act as if they are drawn from some common prior distribution.

Note: alternative assumptions are either that θ_k and σ_k^2 are same for all patients (pooled model) or that they are independent (fixed effects) for each patient

We make the specific distributional assumption that

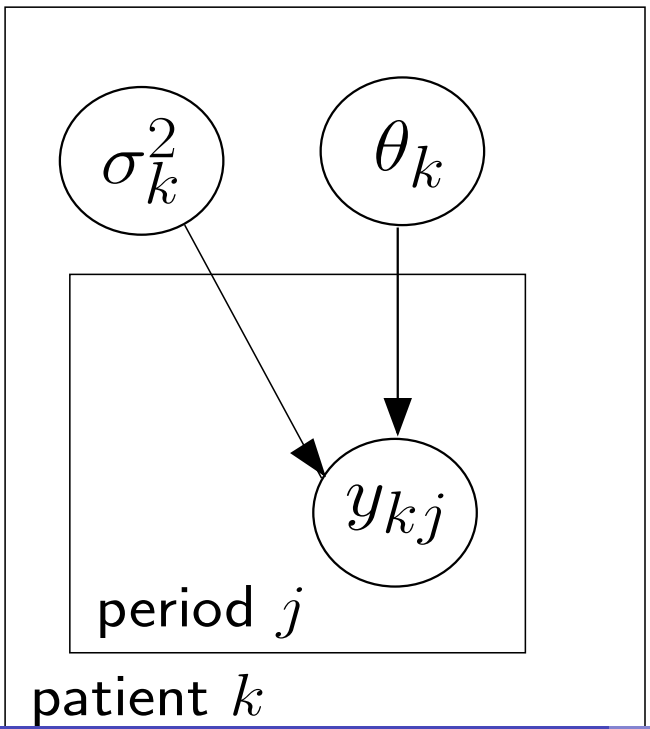
$$\begin{aligned} \theta_k &\sim \mathrm{N}(\mu_\theta, \phi_\theta^2) \\ \log(\sigma_k^2) &\sim \mathrm{N}(\mu_\sigma, \phi_\sigma^2) \end{aligned}$$

A normal distribution for the log-variances is equivalent to a log-normal distribution for the variances

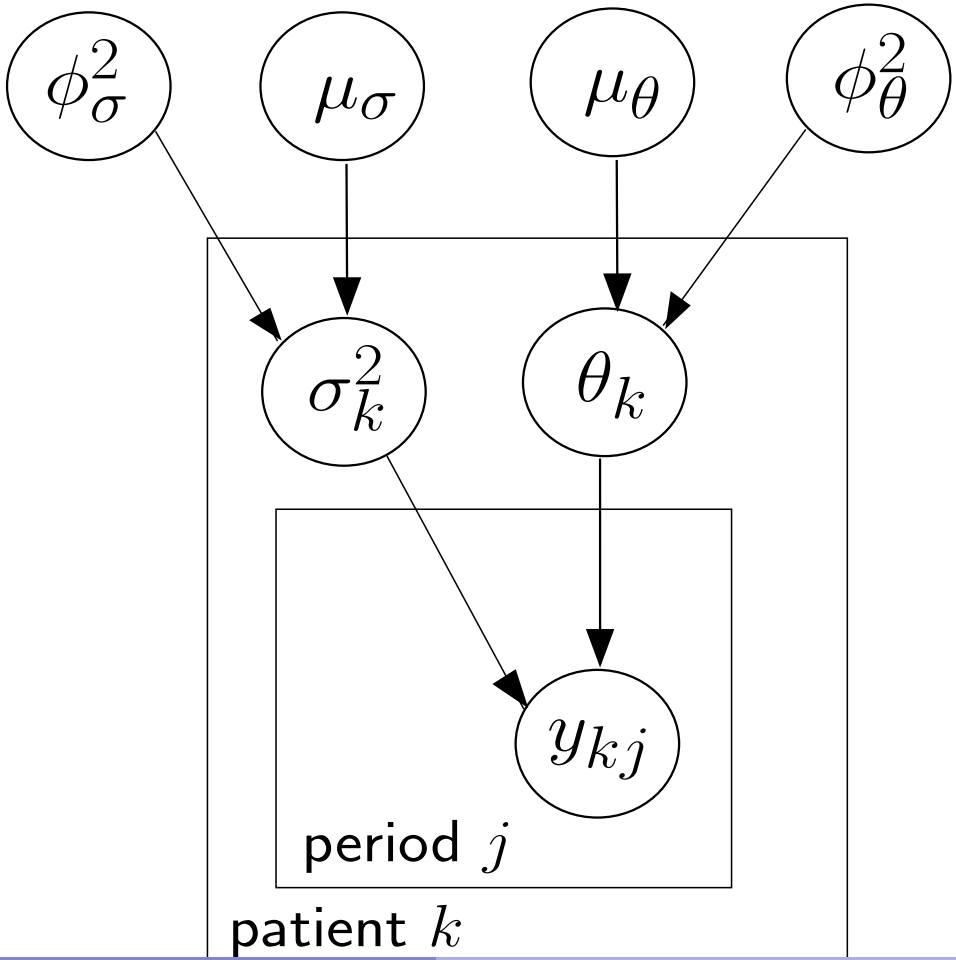
Uniform priors adopted for $\mu_\theta, \phi_\theta, \mu_\sigma$ and ϕ_σ.

Graphical model

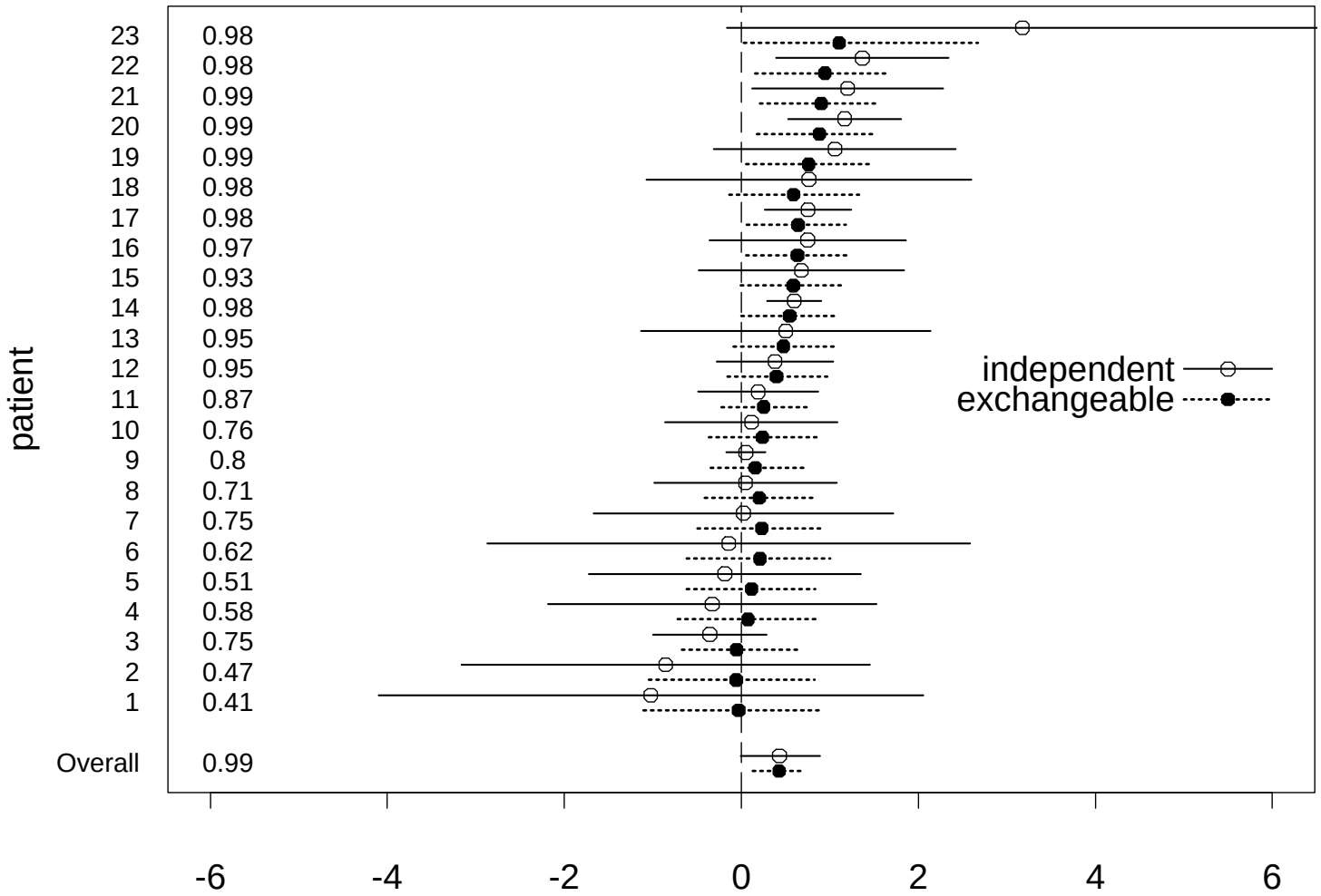
Independent effect



Exchangeable means and variances



Estimates and 95% intervals for treatment effect, and posterior probability that effect > 0



Interpretation

- Exchangeable model shrinks in the extreme patients, reflecting the limited information from each individual (see patient 23)
- Despite shrinkage, narrower intervals mean that 9 patients have 95% intervals excluding 0 compared to 6 with the independent analysis
- One consequence of allowing exchangeable variances is that patient 9 has a *wider* interval under the exchangeable model
  - patient 9's observations were very close together $\rightarrow$ very narrow interval under independence model
- Straightforward to include patient-level covariates
- Sensitivity analysis to the shape of both the sampling and the random-effects distribution: say assuming t-distributions.

Summary

Many reasons to carry out multilevel regression modelling

- Possible to estimate group level regression coefficients, borrowing strength from 'similar' groups $\rightsquigarrow$ reasonable estimates also for very small sample sizes
- Possible to model variation among group level regression coefficients
- Flexible way of treating complex structured data - nested indexes, cross-classifications

Further reading

WinBUGS examples volumes I and II (lots of examples of Bayesian hierarchical models)
Congdon (2001) (lots of examples of Bayesian hierarchical models)
Gelman et al (2004) Chapters 5, 13, 14
Gelman and Hill (2007) Entire book on Regression and hierarchical models in R and WinBUGS - theory and many examples