Session 5. Hierarchical models

Summary

1. Normal hierarchical models
   - Practical Q1
2. Hierarchical regression models
   - Practical Q2
3. Hierarchical models for meta-analysis, and priors in hierarchical models
   - Practical Q3-4

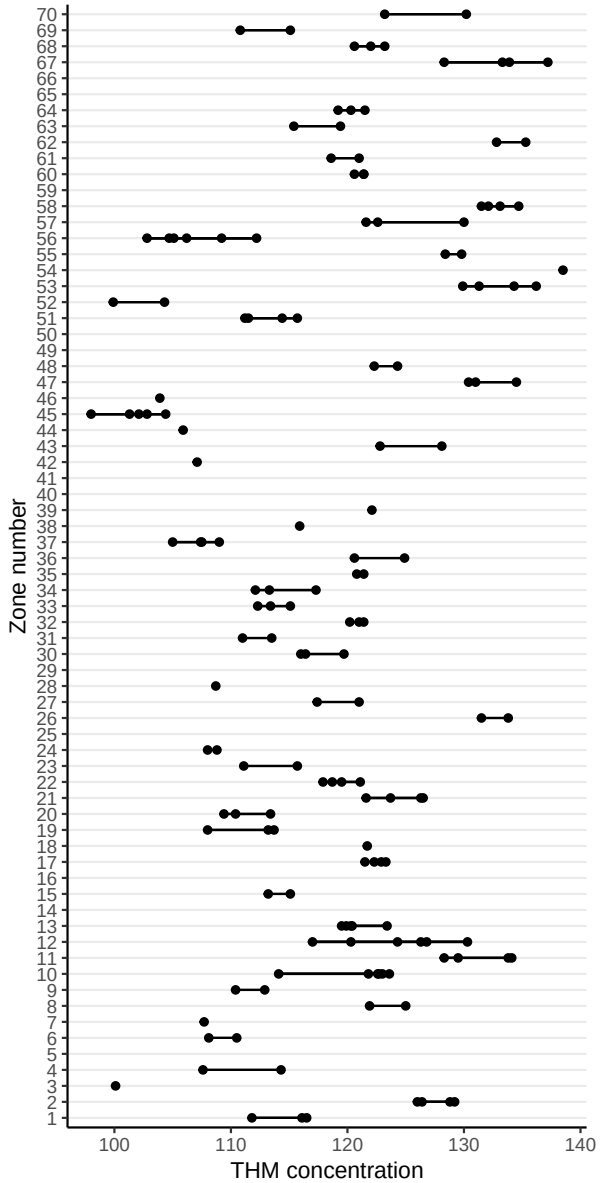
Hierarchical data

Consider data y on concentration of "THM" in water in

- 70 water supply zones: $z = 1, \ldots, 70$
- n_z measurements of THM concentration in each zone z

1. Consider mean THM concentration for zone 70 which had 2 measurements
   - Just take the mean and SE of 2 observations?
   - ...but big uncertainty, since little data
2. Consider a prediction for zone 5 – where no data at all!

Measurements in other zones can give prior information to strengthen inferences



Assumptions for hierarchical data

Make inferences on parameters $\theta_1, \ldots, \theta_k$ measured on k 'units' (individuals, subsets, areas, time-points, trials, etc) which are related by the structure of the problem

Three possible assumptions:

1. **Identical parameters:** All the θ_i are identical
   $\rightarrow$ all the data can be pooled and the individual units ignored.
2. **Independent parameters:** All the θ_i are entirely unrelated
   $\rightarrow$ results from each unit can be analysed independently using a fully specified prior distribution within each unit
   $\rightarrow$ individual estimates of θ_i are likely to be highly variable (unless very large sample sizes)
3. **Exchangeable parameters:** $\theta_1, \ldots, \theta_k$ are related, through being drawn from a common prior distribution with unknown parameters
   $\rightarrow$ a hierarchical model

Example: THM concentrations

- THM measurements for 70 water supply zones included in the data
- We assume a normal likelihood for the data in each zone

$$y_{zi} \sim \mathrm{N}(\theta_z, \sigma^2); \qquad i = 1, \ldots, n_z;\ z = 1, \ldots, 70$$

- Now have 70 distinct mean parameters θ_z
- Take σ^2 to be unknown, with a weakly-informative prior $\sigma \sim \text{Exponential}(\text{rate} = 1/10)$, since expect σ to be around 10 (see Regression lecture)

What prior should we specify for each $\theta_z \ldots$?

Example: THM — identical parameters

- Assume that the mean THM levels are the same in all zones, $\theta_z = \theta$ for all z
- Assign a prior to θ, e.g.

$$\theta \sim \mathrm{N}(120, 100^2)$$

But, assuming $\theta_z = \theta$ is not really sensible since we do not expect different zones to have identical THM levels

Example: THM — independent parameters

- Instead, we might assume independent vague priors for each zone mean, e.g.

$$\theta_z \sim \mathrm{N}(120, 100^2), \qquad z = 1, \ldots, 70$$

- This will give posterior estimates $\mathrm{E}(\theta_z \mid \boldsymbol{y_z}) \approx \overline{\boldsymbol{y_z}}$ (the raw zone mean, which is the MLE)
  - $\rightarrow$ each θ_z estimated independently
  - $\rightarrow$ no 'pooling' or 'borrowing' of information across zones
  - $\rightarrow$ no smoothing of estimates

. . . but are we really including all relevant prior information?

Example: THM — exchangeable parameters

Hierarchical rather than independent priors for each θ_z:

$$\theta_z \sim \mathrm{N}(\mu, \tau^2), \qquad z = 1, \ldots, 70 \qquad \text{(the random effects distribution)}$$

where μ and τ^2 are unknown to also be estimated.

$\rightarrow$ assign priors to μ and τ^2, e.g.

$$\mu \sim \mathrm{N}(120, 100^2)$$
$$\tau \sim \mathrm{Uniform}(0, 100)$$

The quantities μ and τ^2 are called hyperparameters.

Example: THM — exchangeable parameters

$\Rightarrow$ joint prior distribution for the entire set of parameters

$$p(\theta_1, \ldots, \theta_{70}, \sigma^2, \mu, \tau^2) = \left\{ \prod_{z=1}^{70} p(\theta_z \mid \mu, \tau^2) \right\} p(\sigma^2)\ p(\mu)\ p(\tau^2)$$

Then apply Bayes theorem as usual to estimate joint posterior distribution of all the unknown quantities simultaneously:

$$p(\theta_1, \ldots, \theta_{70}, \sigma^2, \mu, \tau^2 \mid \boldsymbol{y})$$
$$\propto \left\{ \prod_{z=1}^{70} \prod_{i=1}^{n_z} p(y_{zi} \mid \theta_z, \sigma^2) \right\} \times \left\{ \prod_{z=1}^{70} p(\theta_z \mid \mu, \tau^2) \right\} \times$$
$$p(\sigma^2) p(\mu)\ p(\tau^2)$$

To get marginal posterior for a particular zone mean parameter $\theta_{z'}$

- integrate the joint posterior $p(\theta_1, \ldots, \theta_{70}, \sigma^2, \mu, \tau^2 \mid \boldsymbol{y})$ over other parameters: σ^2, μ, τ^2, and other θ_zs, $z \neq z'$

Example: THM — exchangeable parameters

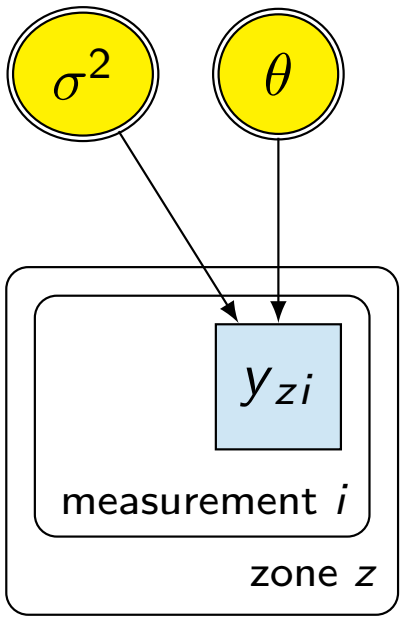
The advantages of this approach are that the posterior for each θ_z

- 'borrows strength' from the likelihood contributions for all of the zones,
  - via the information they provide on population parameters μ and τ^2
- leads to global smoothing or partial pooling of the zone mean THM levels
- reflects our full uncertainty about the true values of μ and τ^2

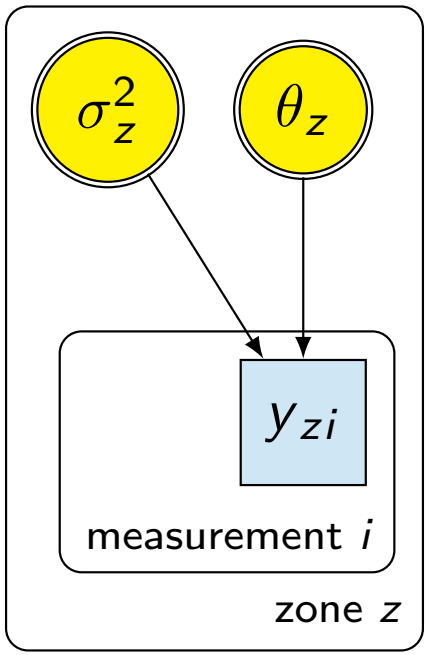
Called Hierarchical, or Random effects, or Mixed effects, or Multilevel models

Example: THM — graphical representation

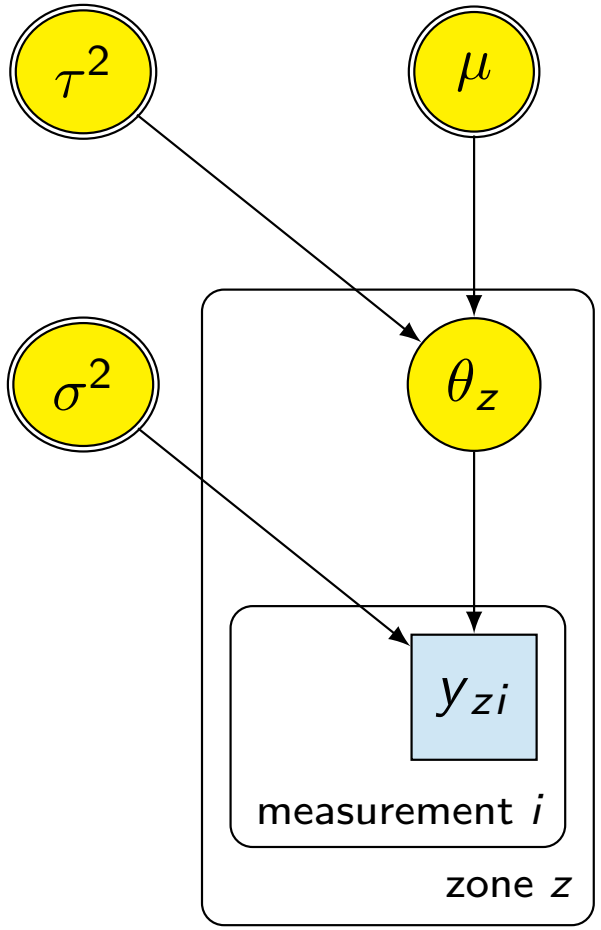
Pooled model



Independent model



Hierarchical model



Example: THM — parameter interpretation

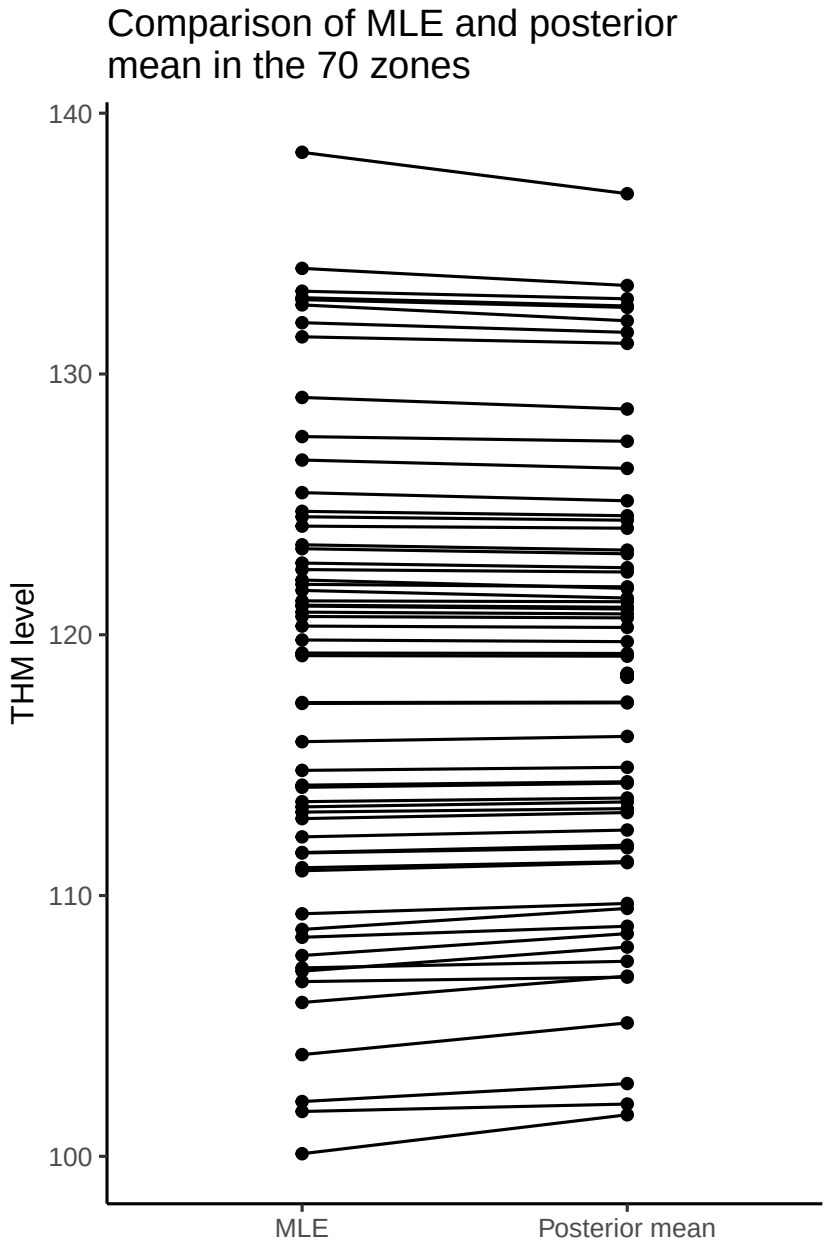
In the hierarchical model:

- θ_z is mean THM concentration in zone z for the study period
  - random effects (of zone on expected THM)
- μ is the overall mean THM concentration across all zones for the study period
- τ^2 is the between-zone variance in THM concentrations
- σ^2 is the residual variance in THM concentrations (reflects measurement error and true within-zone variation in THM levels)

Could similarly allow random variances within zones σ_z^2:

$$
\begin{aligned}
y_{zi} &\sim \mathrm{N}(\theta_z, \sigma_z^2), && i = 1, \ldots, n_z \\
\log\left(\sigma_z^2\right) &\sim \mathrm{N}(\lambda, \nu^2) \\
\lambda &\sim \mathrm{Uniform}(-100, 100) \\
\nu &\sim \mathrm{Uniform}(0, 100)
\end{aligned}
$$

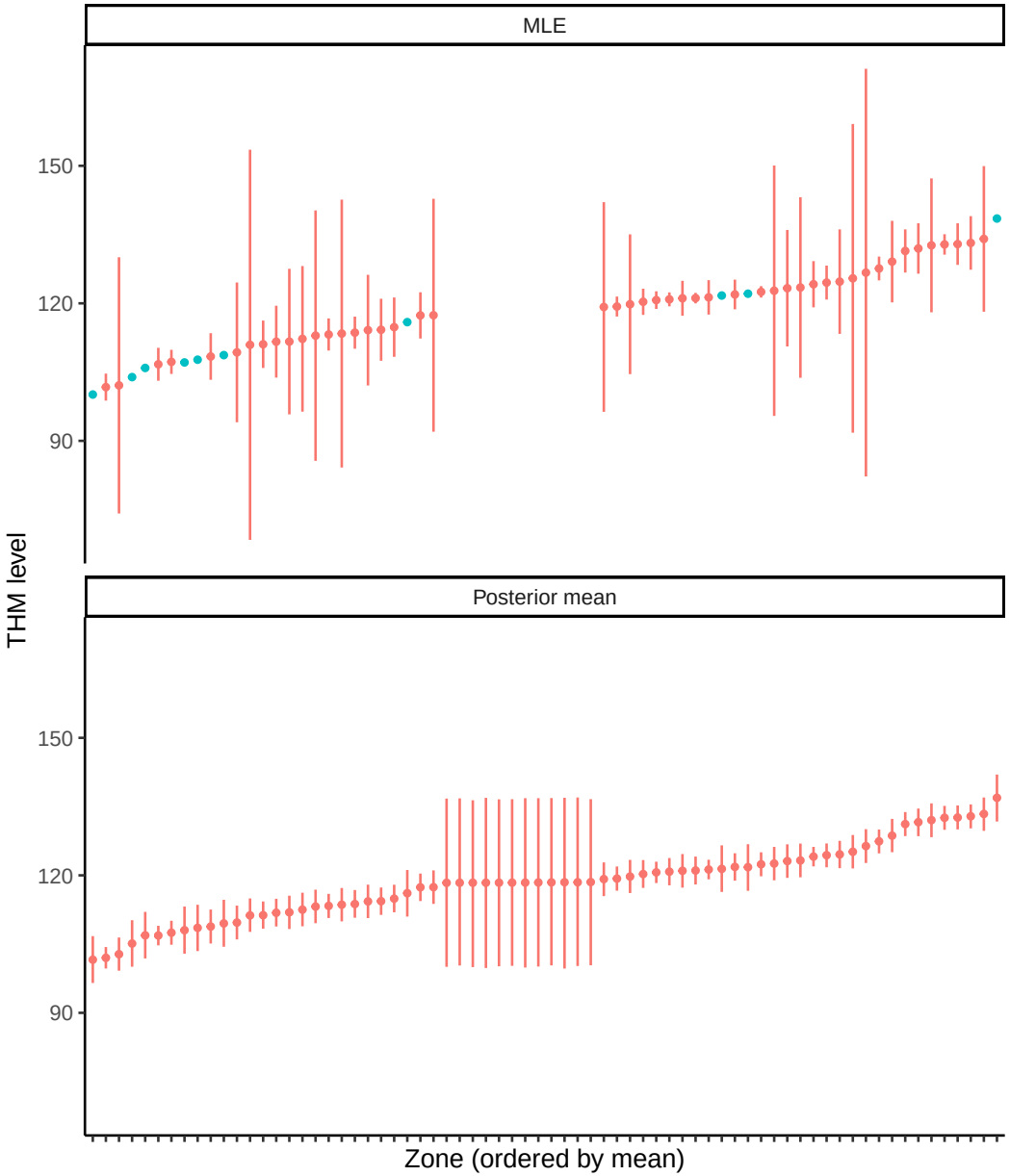
Example: THM — shrinkage of zone means



Posterior mean for a specific area z:

- Weighted average of area-specific mean $\overline{y_z}$, and mean of other areas $\overline{y_{z^c}}$
- Analytic form in this simple normal example: $\frac{\frac{n_z}{\sigma^2}\overline{y_z}+\frac{1}{\tau}\overline{y_{z^c}}}{\frac{n_z}{\sigma^2}+\frac{1}{\tau}}$
- Area-specific mean weighted more highly if more data n_z from that area, or within-area variance σ^2 small compared to between-area variance τ^2

Example: THM — zone mean posterior estimates



- Independent model uses independent zone-specific variances
  - hence no CI for zones with only 1 measurement
- Hierarchical model assumes common error variance
- Hierarchical model also gives estimates for zones with no data!

Example: THM — data format for hierarchical data

Data contain between 0 and 6 observations per zone → 'ragged array'

Zone	THM level
1	111.3, 112.9, 112.9, 105.5
2	122.6, 124.6, 135.4, 135.7, 156.7, 144.8
3	133.1, 116.6, 106.2, 126
4	111.6, 112.5, 98.6, 107.7
5	–
6	124.7
..	....

How to code model and data in JAGS?

Example: THM — data format for JAGS

Associate each observation with the zone where it came from in the data:

```
list(
  n = 173,
  n_zone = 70,
  thm = c(111.3, 112.9, 112.9, 105.5, 122.6, 124.6,
          135.4, 135.7, 156.7, 144.8, 133.1, 116.6,
          106.2, 126, 111.6, 112.5, 98.6, 107.7,
          124.7, ...),
  zone = c(1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3,
           4, 4, 4, 4, 6, ...))
```

Example: THM – model specification in JAGS

```
for (i in 1:n){
  thm[i] ~ dnorm(theta[zone[i]], sigma.prec) # likelihood
}
for (z in 1:n_zone) {
  # zone-specific means (random effects)
  theta[z] ~ dnorm(mu, tau.prec)
}

mu ~ dnorm(120, 100^2) # prior random effects mean
# random effects SD (between-zone SD of mean THM)
tau.sd ~ dunif(0, 100)
tau.var <- pow(tau.sd, 2)
tau.prec <- 1 / tau.var

sigma.sd ~ dexp(1/10)
sigma.prec <- 1/sigma.sd^2
```

Variance Partition Coefficient (VPC)

The residual variation in the response variable is split into components attributed to different levels in hierarchical or multilevel models

Often of interest to quantify percentage of total variation attributable to higher level units e.g. groups vs individuals

- We can use VPC, or intra-cluster correlation coefficient (ICC)

$$\text{VPC} = \frac{\tau^2}{\tau^2 + \sigma^2}$$

  - σ^2 is the 'within-group' variance (i.e. of Normal likelihood)
  - τ^2 is the 'between-group' variance (i.e. of random effects)
- To calculate VPC, add the calculation to the model (see practical)

General form for hierarchical model with m levels

In general, suppose we have data $\boldsymbol{y}$ and parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$

- Likelihood $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ models structure of the observables
- Prior $p(\boldsymbol{\theta})$ is decomposed into conditional distributions $p(\boldsymbol{\theta} \mid \phi_2) \times p(\phi_2 \mid \phi_3) \times \ldots \times p(\phi_m)$

The marginal prior distribution for $\boldsymbol{\theta}$ is then

$$p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi_2)\, p(\phi_2 \mid \phi_3) \ldots p(\phi_{m-1} \mid \phi_m)\, p(\phi_m)\, d\phi_2\, d\phi_3 \ldots d\phi_m$$

- ϕ_ℓ are called hyperparameters of level ℓ and are introduced to simplify prior specification.
- The conditional prior distributions $p(\phi_{\ell-1} \mid \phi_\ell)$ express structural judgements (exchangeability)
- Theoretically there can be as many levels as necessary, but can be hard to interpret parameters if too many

Non-Bayesian methods for hierarchical modelling

Likelihood $p(\boldsymbol{y} \mid \boldsymbol{\theta})$, random effects distribution $p(\boldsymbol{\theta} \mid \phi)$, hyperparameters ϕ

Maximum (restricted) marginal likelihood: integrate out random effects θ and then maximise $p(\boldsymbol{y} \mid \phi)$

$\rightarrow$ maximum likelihood estimates of ϕ and any variance components

- e.g. using (R)IGLS = (Restricted) Iterative Generalised Least Squares for normal models, or numerical (Gauss-Hermite) integration more generally
- implemented e.g. in R (`lme`, `nlme`), Stata (`mixed`) SAS (`proc mixed`), MLWin

No prior distribution is assumed for ϕ.

- Ignoring full uncertainty in ϕ can lead to overly precise estimates
- Computationally easier than MCMC for standard models
- Less generalisable to more complex data structures

Practical 1

- THM example
  - Learn how to code hierarchical models in JAGS (and prepare the data in R)
  - Identify and interpret the parameters of a hierarchical model
  - Compare MLEs and posterior means (as an example of working with results of a hierarchical model)
  - Adapt an existing hierarchical model in JAGS (hierarchical variance)

Assumptions in hierarchical models

Hierarchical models explicitly model all sources of variation, within and between groups but any model has assumptions

Choice of random effects distribution? Any distribution possible with MCMC.

- e.g. $t-$distribution, heavier tails than Normal, more robust to outlying units

Assumption of exchangeability between groups?

- Exchangeability assumes all group are "equivalent apriori"
- Covariates may explain systematic differences between groups
- After adjusting for covariates, groups might then be assumed exchangeable on the population parameters

Extend hierarchical models to include regression on covariates

Example: Hep B — hierarchical regression models

- Program of childhood vaccination against Hepatitis B (HB) in Gambia
- Program effectiveness depends on duration of immunity afforded by vaccination

Data:

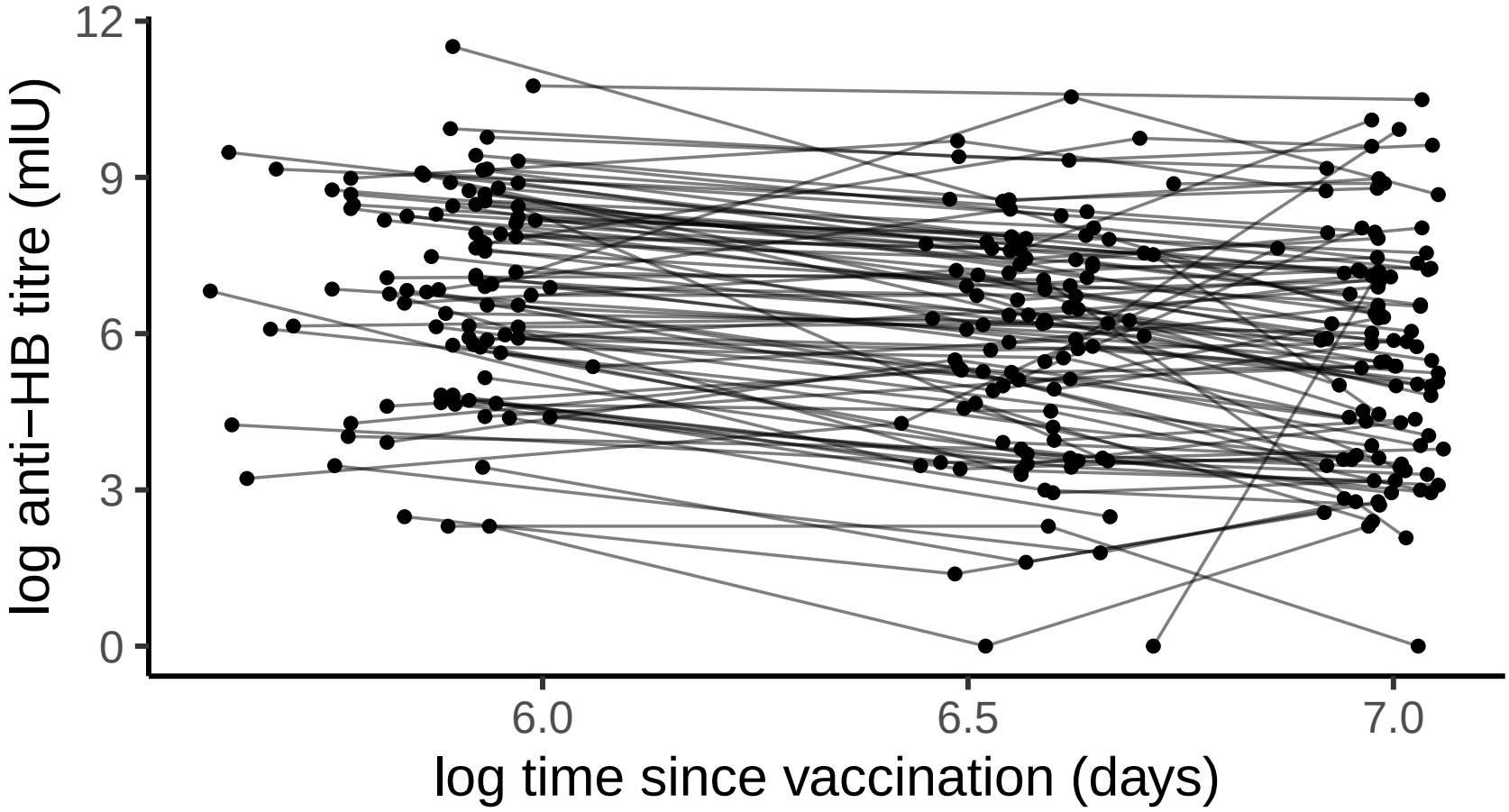
- 106 children immunized against HB
- Anti-HB titre measured at time of vaccination (baseline) and on 2 or 3 follow-up occasions

Build model to predict

- an individual child's protection against HB
- as a function of time.

Example: Hep B — raw data

A similar study found anti-HB titre $\propto$ time since vaccination^{-1}, which suggests log-anti-HB-titre and log-time-since-vaccination should be approximately linear



Example: Hep B — basic linear model (non-hierarchical)

1. Likelihood for responses y_{ij} = log of the jth anti-HB titre measurement for child i

$$y_{ij} \sim \mathrm{N}(\mu_{ij}, \sigma^2)$$

2. Linear predictor:

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

where

t_{ij} = log 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 intercept α, and slope β for all children
- Ignores potential correlation between repeated measurements from same child

Example: Hep B — hierarchical linear model

Modify model to allow separate intercept and slope for each child:

$$y_{ij} \sim \mathrm{N}(\mu_{ij}, \sigma^2)$$
$$\mu_{ij} = \alpha_i + \beta_i(t_{ij} - \overline{t}) + \gamma(y_{0i} - \overline{y_0})$$

- Assumes the conditionally on α_i and β_i, $\{y_{ij}, j = 1, 2, ..\}$ are independent
- Assume all α_i are exchangeable and all β_i are exchangeable

e.g.
$$\alpha_i \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2) \quad i = 1, \ldots, 106$$
$$\beta_i \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2) \quad i = 1, \ldots, 106$$

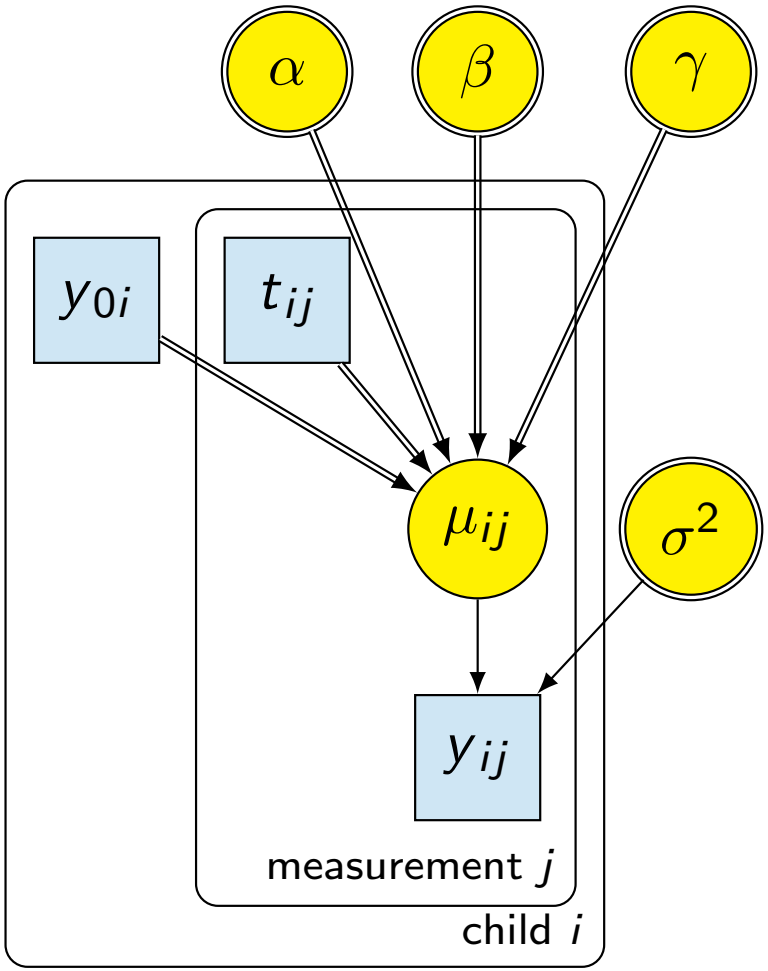
- Need priors for the hyperparameters of the population distribution, e.g. if we wanted to be vague, could choose

$$\mu_\beta, \mu_\alpha \sim \mathrm{N}(0, 100^2) \qquad \sigma_\alpha^2, \sigma_\beta^2 \sim \mathrm{Uniform}(0, 100)$$

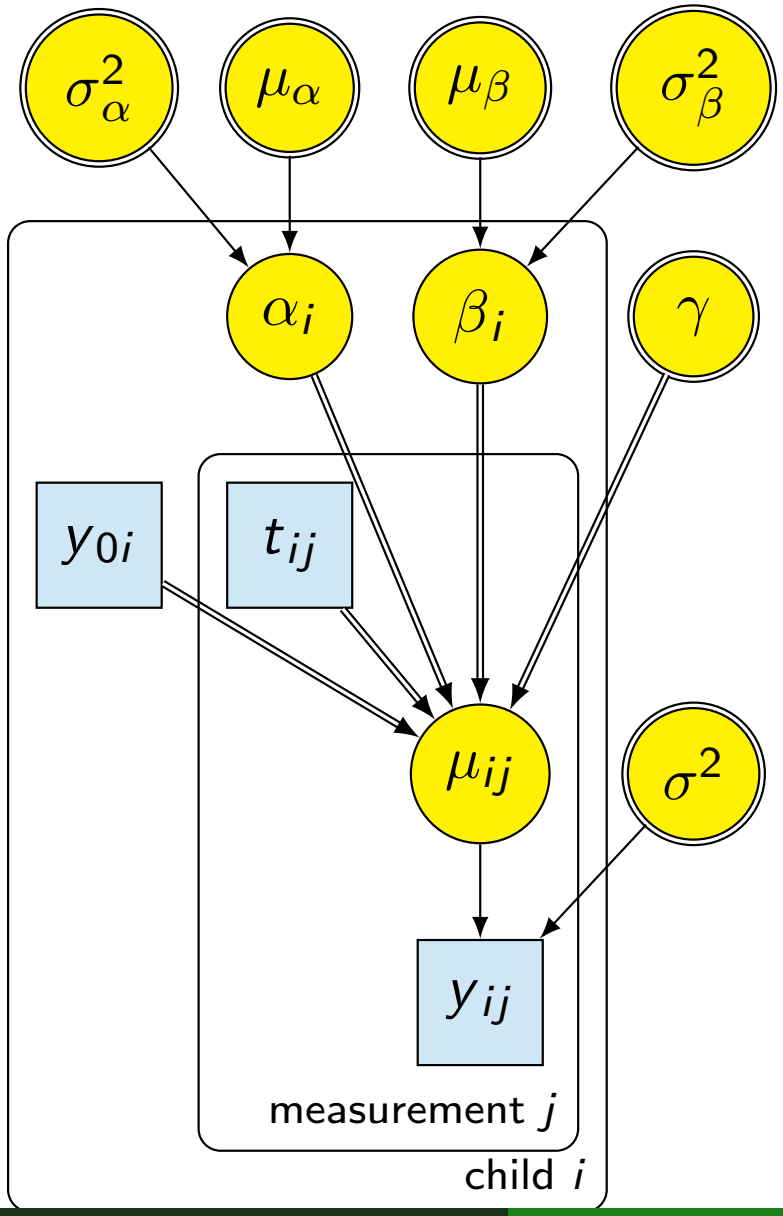
Called a Hierarchical Linear Model, Linear Mixed Model or Random Coefficients model

Example: Hep B — graph representation

Linear model



Hierarchical linear model



Example: Hep B — model specification in JAGS

```
for (i in 1:n) {
  log_anti_hb[i] ~ dnorm(mu[i], sigma.prec) # likelihood

  mu[i] <- alpha[child[i]] +
           beta[child[i]] * log_time_centred[i] +
           gamma * log_anti_hb_at_time0_centred[i]
}
for (c in 1:n_child){
  # Random effects distributions for alpha and beta
  alpha[c] ~ dnorm(mu.alpha, sigma.alpha.prec)
  beta[c] ~ dnorm(mu.beta, sigma.beta.prec)
}
# ...plus priors for mu.alpha, sigma.alpha.prec,
# mu.beta, sigma.beta.prec, sigma.prec, gamma
```

Example: Hep B — results

Parameters	Simple linear	Parameters	Hierarchical
α	6.14 (0.10)	μ_α	6.15 (0.15)
		σ^2_α	1.42 (0.12)
β	-1.05 (0.22)	μ_β	-1.06 (0.14)
		σ^2_β	0.32 (0.21)
γ	0.67 (0.06)	γ	0.67 (0.08)
σ^2	1.73 (0.07)	σ^2	1.00 (0.06)

- Means from hierarchical model are similar to the simple linear model
- Residual variance σ^2 has been considerably reduced: more variation between children (σ^2_α) than within (σ^2)

Practical 2

- Hep B example of hierarchical regression
  - Adapt the non-hierarchical regression model into the hierarchical model

Meta analysis

Meta analysis is a method for summarising results from a number of separate studies of treatments or interventions.

Forms part of the process of systematic review, which also includes the process of identifying and assessing the quality of relevant studies.

- Meta-analysis is very widely adopted in medical applications
- Each 'study' is often a single clinical trial
- Can be viewed as a special case of hierarchical modelling.

Meta analysis for odds ratios

Often we want to compare the number of 'successes' or 'events' between patients who were treated and those in a 'control' group (arm)

	Treatment		Control	
Study	Successes	Total	Successes	Total
Study 1	y_{1T}	n_{1T}	y_{1C}	n_{1C}
Study 2	y_{2T}	n_{2T}	y_{2C}	n_{2C}
⋮	⋮	⋮	⋮	⋮
Study s	y_{sT}	n_{sT}	y_{sC}	n_{sC}
⋮	⋮	⋮	⋮	⋮

The number of 'failures' in study s are

- Treatment group $n_{sT} - y_{sT}$
- Control group $n_{sC} - y_{sC}$

Compare the number of successes using the (log) odds ratio

Common effect meta-analysis for odds ratios

Assume a single underlying treatment effect δ common to all studies

For studies $s = 1, \ldots, N$

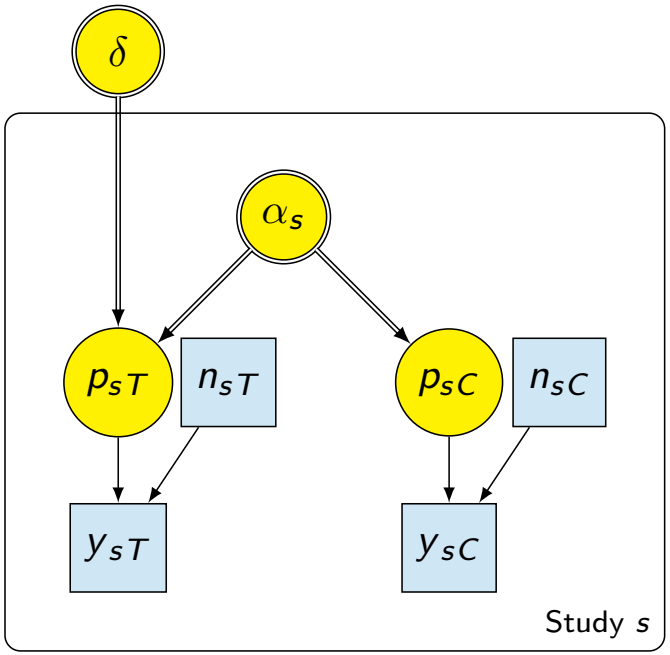
$$y_{sT} \sim \text{Binomial}(n_{sT}, p_{sT}) \qquad y_{sC} \sim \text{Binomial}(n_{sC}, p_{sC})$$
$$\text{logit}(p_{sT}) = \alpha_s + \delta \qquad \text{logit}(p_{sC}) = \alpha_s$$

- α_s: log odds for a 'success' in control group in study s
- $\alpha_s + \delta$ is log odds for 'success' in treatment arm in study s
- δ is the log odds ratio

This is an identical parameter model.

Common effect meta-analysis is also called fixed-effects meta-analysis.

Common effect meta-analysis for odds ratios




```
for (s in 1:Ns){
  for (a in 1:2){
    y[s,a] ~ dbin(p[s,a], n[s,a])
    logit(p[s,a]) <- alpha[s] +
                     delta[a]
  }
  alpha[s] ~ dnorm(0, 1/10^2)
}
delta[1] <- 0
delta[2] ~ dnorm(0, 1/10^2)
```

Random effects meta-analysis for odds ratios

Procedures, patient characteristics etc often differ between studies, and often at least part of the variability is unexplained

For studies $s = 1, \ldots, N$

$$y_{sT} \sim \text{Binomial}(n_{sT}, p_{sT}) \qquad y_{sC} \sim \text{Binomial}(n_{sC}, p_{sC})$$

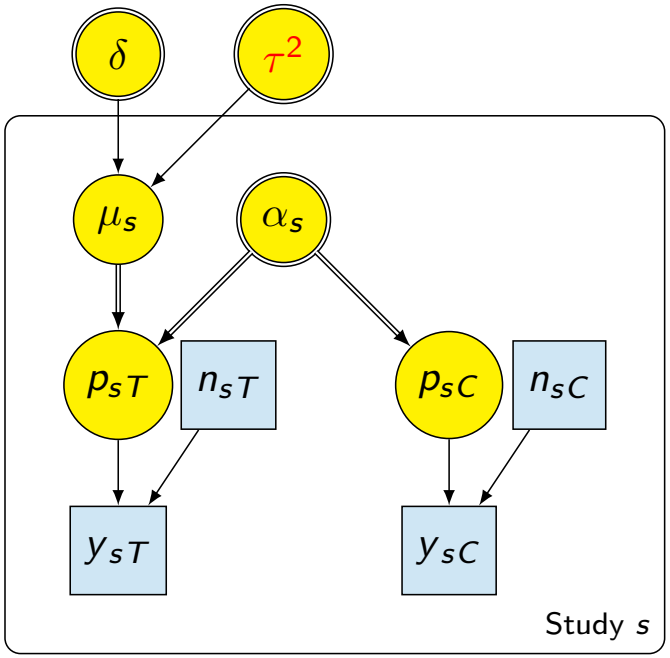
$$\text{logit}(p_{sT}) = \alpha_s + \mu_s \qquad \text{logit}(p_{sC}) = \alpha_s$$

$$\mu_s \sim N(\delta, \tau^2)$$

- α_s is the log odds for a 'success' in control arm in study s
- $\alpha_s + \mu_s$ is the log odds for a 'success' in treatment arm in study s
- δ is the overall log odds ratio

This is an exchangeable parameter model.

Random effects meta-analysis for odds ratios




```
for (s in 1:Ns){
  for (a in 1:2){
    y[s,a] ~ dbin(p[s,a], n[s,a])
    logit(p[s,a]) <- alpha[s] +
                          mu[s,a]
  }
  mu[s, 1] <- 0
  mu[s, 2] ~ dnorm(delta, prec.mu)

  alpha[s] ~ dnorm(0, 1/10^2)
}
delta ~ dnorm(0, 1/10^2)

prec.mu <- 1/(sd.mu * sd.mu)
sd.mu ~ dunif(0, 10)
```

Example: Sepsis (Ohlsson and Lacy, 2013)

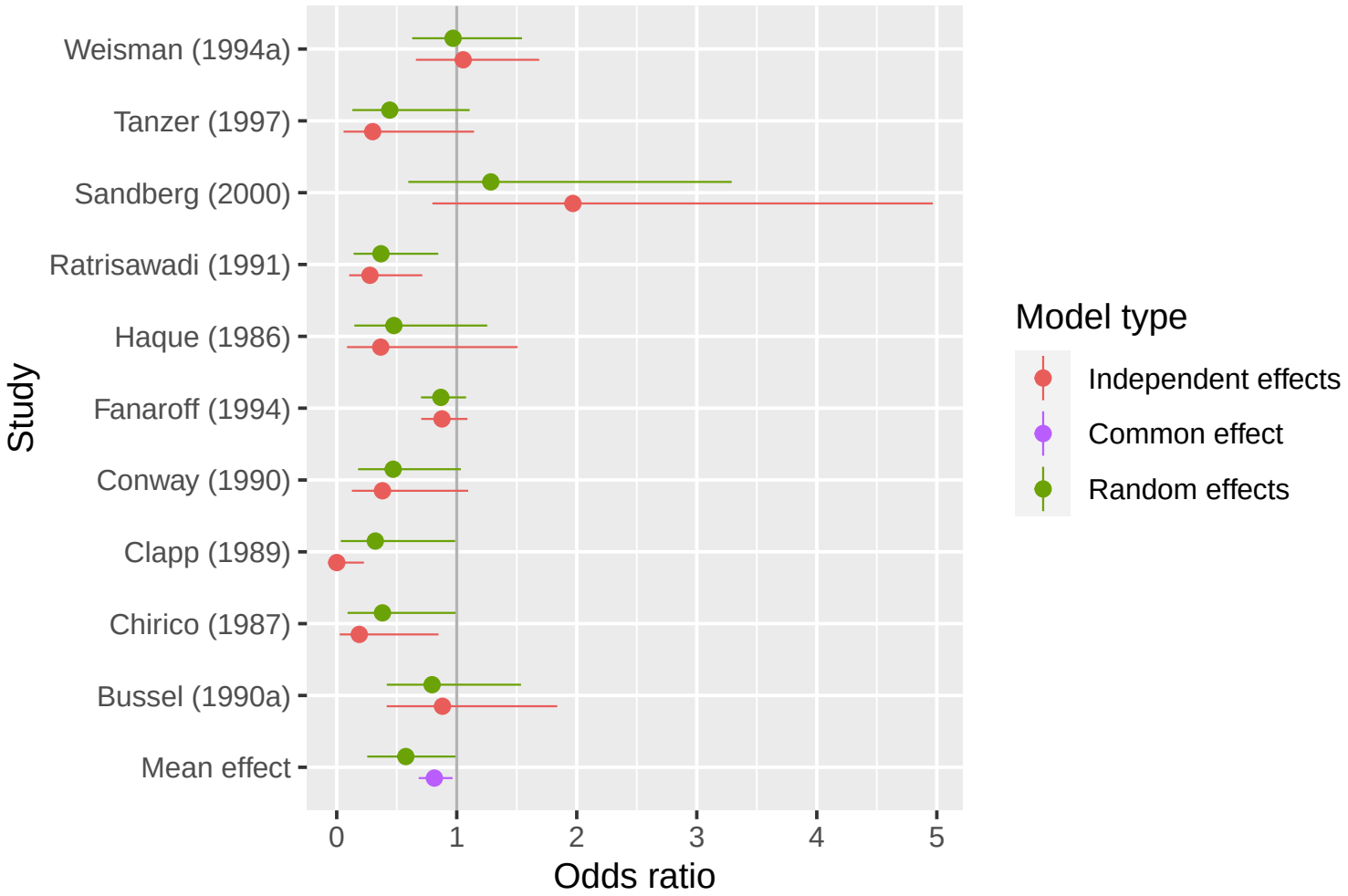
Outcome Infection (or not) in preterm/low birth weight infants

Arms Intravenous immunoglobulin (IVIG) vs placebo

Question Does administration of IVIG prevent infection in hospital, compared to placebo? Event = 'sepsis'

	Treatment		Control	
Study	Events	Total	Events	Total
Bussel (1990a)	20	61	23	65
Chirico (1987)	2	43	8	43
Clapp (1989)	0	56	5	59
Conway (1990)	8	34	14	32
Fanaroff (1994)	186	1204	209	1212
Haque (1986)	4	100	5	50
Ratrisawadi (1991)	10	68	13	34
Sandberg (2000)	19	40	13	41
Tanzer (1997)	3	40	8	40
Weisman (1994a)	40	372	39	381

Example: Sepsis — results



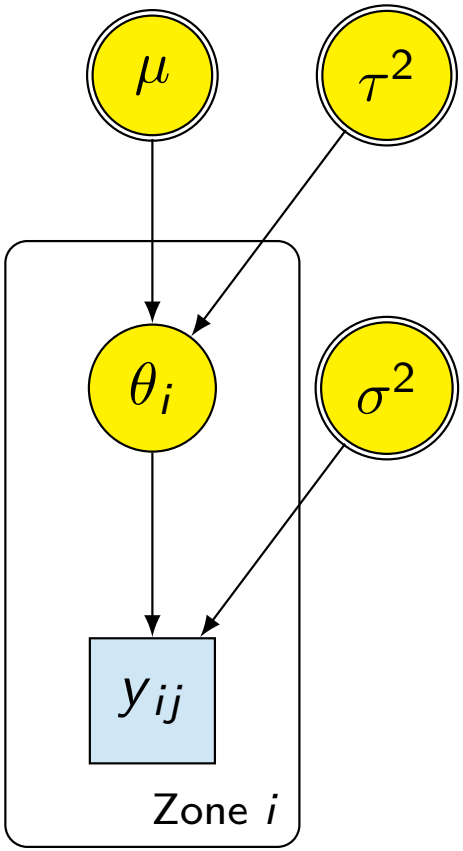
Meta-analysis summary

Hierarchical data asssumption	Meta analysis model
Identical parameters	Common efffect meta-analysis
Assume the parameters are identical, so just pool all the data/studies together	
Independent parameters	Study-specific models/raw data
Assume the parameters are unrelated, so analyse each study completely separately	
Exchangeable parameters	Random effects meta-analysis
Assume the parameters are 'similar' (or 'exchangeable'), use a hierarchical model	

Priors for hierarchical models

Normal-normal hierarchical model

$$y_{ij} \sim \mathrm{N}(\theta_i, \sigma^2)$$
$$\theta_i \sim \mathrm{N}(\mu, \tau^2)$$



Three parameters that need prior distributions

1. σ^2 – variance of residual
2. μ – the population mean
3. τ^2 – variance of unit-specific parameters

Usually the choice of prior for (1) and (2) is not crucial (the data are usually informative enough)

But (3) requires considerable care

Priors — mean and residual variance

Standard choices are as per the Regression lecture.

Residual/error variance σ^2

- Vague priors: e.g. $1/\sigma^2 \sim \text{Gamma}(\epsilon, \epsilon)$, or $\log(\sigma) \sim \text{Uniform}(0, L)$, for some L
- Weakly informative priors e.g. $\sigma \sim \text{Exponential}(\dots)$

Population mean μ

- Should represent our belief about the average value in our population (NOT the distribution of values we expect to see – recall Regression lecture)
- Vague priors. e.g. $\mu \sim \text{N}(0, 100^2)$
- Informative priors e.g. $\mu \sim \text{Uniform}(a, b)$; or specify quantiles, then identify a distribution with those quantiles

Priors — random effect variances — vague priors

Random effect variances are difficult, because between-unit variation could often be alternatively explain by large within-unit variation.

There is no vague prior that will always be best in all settings.

- A widely recommended choice is a uniform prior on the standard deviation scale, e.g.

$$\tau \sim \text{Uniform}(0, 100)$$

  This is a proper prior that is an approximation to the improper prior $\tau \propto 1$
- Alternatively, Gelman (Bayesian Analysis, 2006) recommends a half-t or half-normal distribution on the standard deviation scale

Priors: random effect variances — informative priors

Meta-analysis is often one of the trickiest cases – there are often very few trials.

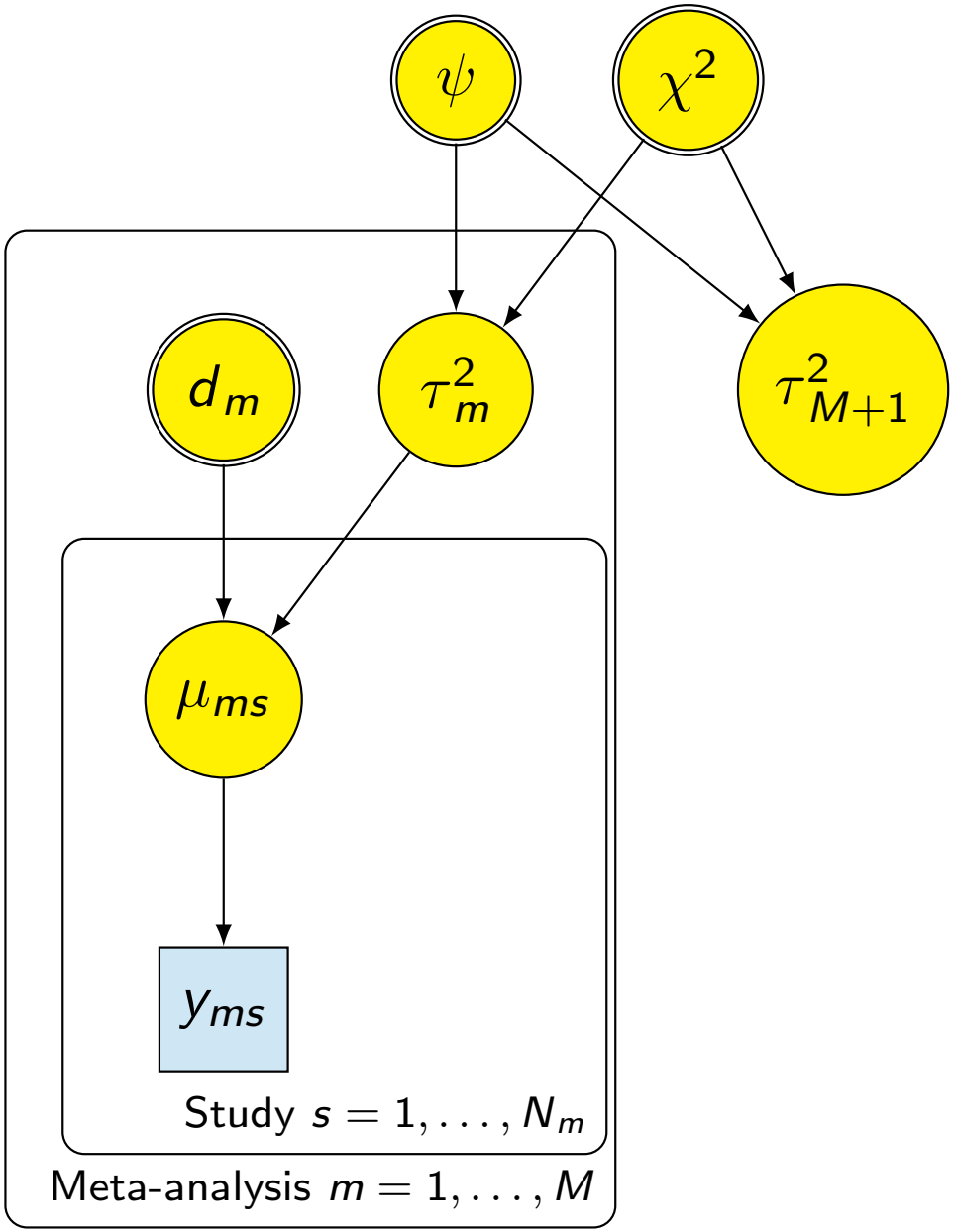
Can we use previous meta-analyses to construct a reasonable prior for the random-effect variance?

Turner *et al.* (2012) used data from over 14,000 meta-analyses from Cochrane Database of Systematic Reviews

1. Start with the standard random effects meta-analysis model for each meta-analysis
2. Add an additional level to the model, describing variation in τ^2
3. Obtain predictive distribution for τ^2 for a new meta-analysis

This predictive distribution can be used as an informative prior

Priors: random effect variances — informative priors model



For study $s = 1, \ldots, N_m$ within meta-analysis $m = 1, \ldots, M$ assume random-effect meta analysis model as before

$$y_{ms} \sim \text{as per slide 35}$$

$$\mu_{ms} \sim \text{N}(d_m, \tau^2_m)$$

$$\log(\tau^2_m) \sim \text{N}(\psi, \chi^2)$$

For a new meta-analysis ($m = M + 1$), we can predict τ^2_{M+1}:

$$\log(\tau^2_{M+1}) \sim \text{N}(\psi, \chi^2)$$

Priors: random effect variances — informative example

Some types of outcomes and comparisions may be more heterogenous than others.

So Turner et al added a regression model component, and provided suggested informative priors for τ^2 in each case:

	Pharmacological vs Placebo/Control	Pharmacological vs Pharmacological
All-cause mortality	Log-normal $(-4.06, 1.45^2)$: median = 0.017; 95% range = (0.001–0.30)	Log-normal $(-4.27, 1.48^2)$: median = 0.014; 95% range = (0.0008–0.25)
Semi-objective[b]	Log-normal $(-3.02, 1.85^2)$: median = 0.049; 95% range = (0.001–1.83)	Log-normal $(-3.23, 1.88^2)$: median = 0.040; 95% range = (0.001–1.58)

Practical 3

1. Sepsis example
   - Interpreting the parameters of a meta-analysis model
   - Adding a regression component to a meta-analysis model
2. Prior sensitivity in hierarchical models
   - Formulating an informative prior
   - Comparing posteriors under different priors

Hierarchical models: key concepts

1. **Explicitly model all sources of variation** within and between groups. Can construct hierarchical models for means, for regression coefficients, variances, etc.

2. **Exchangeability**: unit-specific parameters assumed to be similar, through being drawn from a common population distribution with unknown mean/variance

3. **Borrowing of strength**: information about one unit is supplemented by information from the population distribution of values.

4. **Shrinkage**: outlying unit-specific estimates tend to be pulled towards the population means. Units with less data are shrunk more.