

7. Aggregated level data

Gianluca Baio

Department of Statistical Science | University College London

- ✉ g.baio@ucl.ac.uk
- 🌐 <https://gianluca.statistica.it/>
- 🌐 <https://egon.stats.ucl.ac.uk/research/statistics-health-economics/>
- 🌐 <https://github.com/giabaio>
- 🌐 <https://github.com/StatisticsHealthEconomics>
- 🐦 [@gianlubaio](https://twitter.com/gianlubaio)

STAT0019 - Bayesian Methods in Health Economics, UCL

- Role of evidence synthesis in decision modelling
 - Absolute and relative effects
- Meta-analysis of aggregated summaries from RCTs
 - Multi-level hierarchical models
 - Exchangeability: No/Partial/Complete pooling
 - Magnesium example
- Evidence synthesis in health economics
 - Influenza example

References

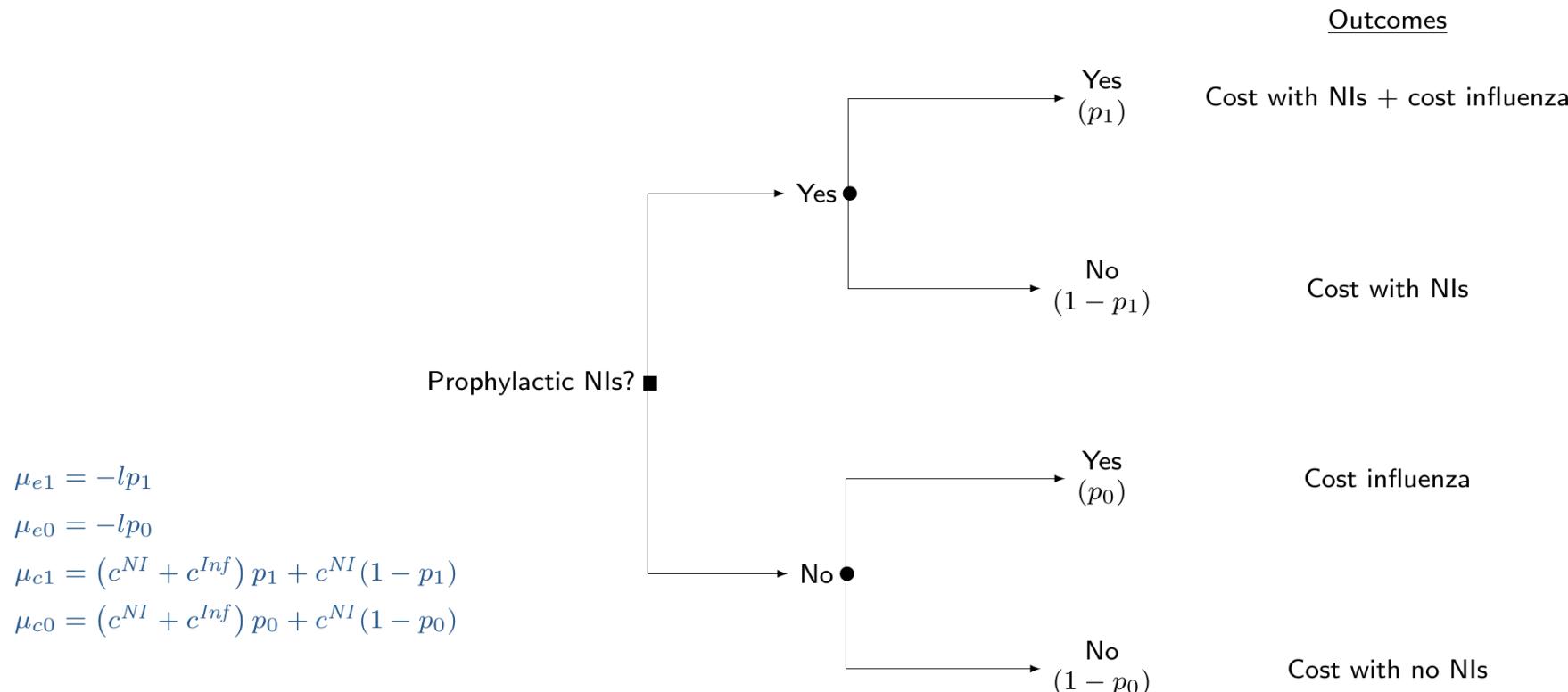
- *The BUGS Book*, chapter 10  Library  Book website
- *Bayesian Methods in Health Economics*, chapter 5.3  Library  Book website (CRC)  Book website  Code
- NICE DSU Evidence Synthesis Technical Support Document Series
- *Evidence Synthesis for Decision Making in Healthcare*  Library  Book website
- *Data Analysis Using Regression and Multilevel/Hierarchical Models*  Library  Book website (CUP)  Book website

- Robust healthcare decisions should:
 - Consider all costs and benefits over a patients lifetime
 - Be evidence-based, reflecting all relevant evidence
- Economic evaluation based on individual level data from a single RCT
 - May not capture longer-term costs/benefits
 - May not generalise to other populations/settings
 - May not be the only relevant source of evidence... results may differ if we use another source

- Robust healthcare decisions should:
 - Consider all costs and benefits over a patients lifetime
 - Be evidence-based, reflecting all relevant evidence
- Economic evaluation based on individual level data from a single RCT
 - May not capture longer-term costs/benefits
 - May not generalise to other populations/settings
 - May not be the only relevant source of evidence... results may differ if we use another source
- **Need to identify all relevant evidence**
 - Transparent, reproducible, specific
- Systematic review
 - Population (including subgroups)
 - Interventions
 - Comparators
 - Outcomes

("Standard") Statistical modelling

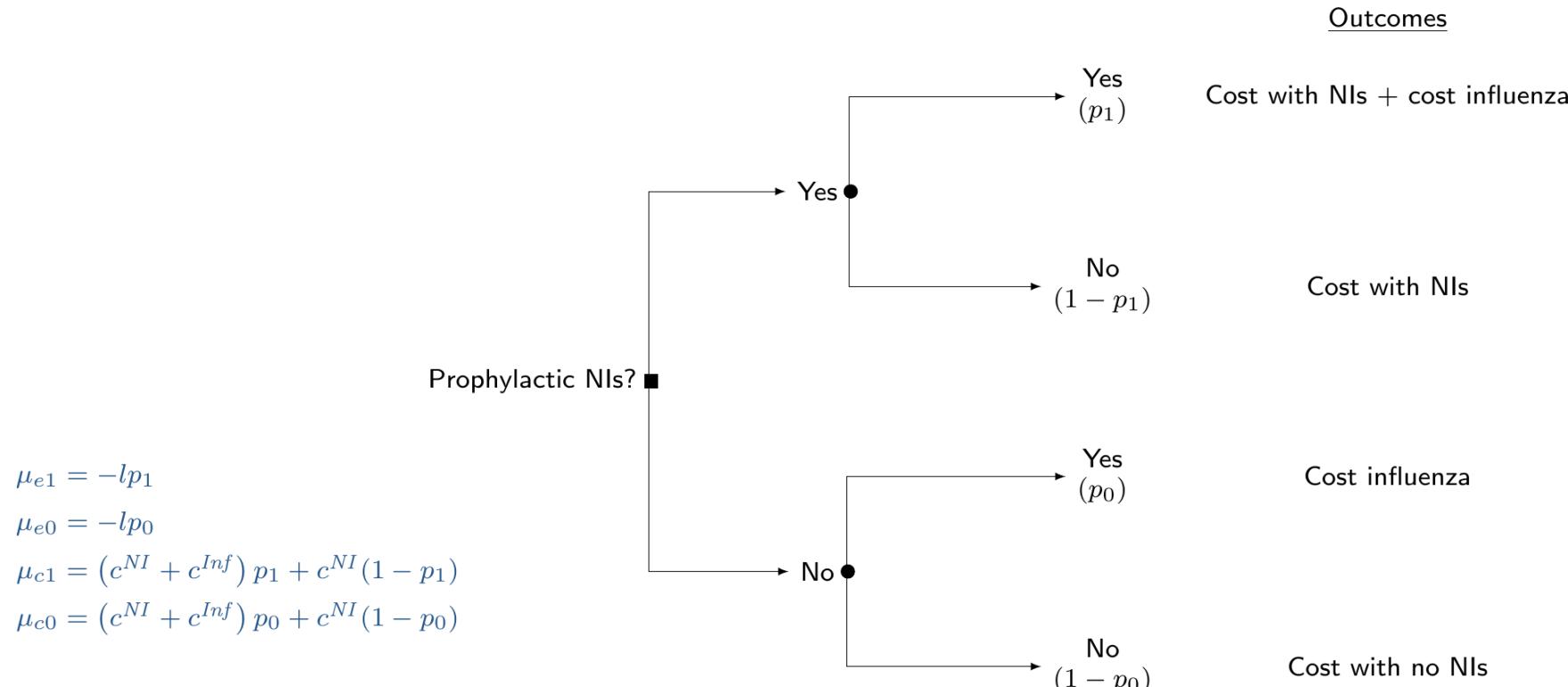
- Build a population level model (eg decision tree/Markov model)



NB: in this case, the "data" are typically represented by summary statistics for the parameters of interest $\theta = (p_0, p_1, l, \dots)$, but may also have access to a combination of ILD and summaries

("Standard") Statistical modelling

- Build a population level model (eg decision tree/Markov model)



NB: in this case, the "data" are typically represented by summary statistics for the parameters of interest $\theta = (p_0, p_1, l, \dots)$, but may also have access to a combination of ILD and summaries

- Use point estimates for the parameters to build the “base-case” (average) evaluation
- Use resampling methods (eg bootstrap) to propagate uncertainty in the point estimates and perform uncertainty analysis

What's wrong with this?...

- Potential correlation between costs & clinical benefits [Individual level + Aggregated level Data]
 - Strong positive correlation - effective treatments are innovative and result from intensive and lengthy research ⇒ are associated with higher unit costs
 - Negative correlation - more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
 - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation - MCMC does a better job!
- Joint/marginal normality not realistic [Mainly ILD]
 - Costs usually skewed and benefits may be bounded in $[0; 1]$
 - Can use transformation (e.g. logs) - but care is needed when back transforming to the natural scale
 - Should use more suitable models (e.g. Beta, Gamma or log-Normal) - generally easier under a Bayesian framework
 - Particularly relevant in presence of partially observed data - more on this later!
- Particularly as the focus is on decision-making (rather than just inference), we need to use **all available evidence** to fully characterise current uncertainty on the model parameters and outcomes Mainly ALD
 - A Bayesian approach is helpful in combining different sources of information
 - **Propagating uncertainty is a fundamentally Bayesian operation!**

- **Absolute** effects
 - e.g. probabilities, mean scores, event rates
 - Typically what's needed in a decision model
- **Relative** effects
 - e.g. log-odds ratio, mean difference, hazard ratio
 - By design RCTs provide evidence on relative effects
 - . . . relative effects more generalisable than the absolute effects

- **Absolute** effects
 - e.g. probabilities, mean scores, event rates
 - Typically what's needed in a decision model
- **Relative** effects
 - e.g. log-odds ratio, mean difference, hazard ratio
 - By design RCTs provide evidence on relative effects
 - . . . relative effects more generalisable than the absolute effects

Goal: Apply relative effects to reference absolute effect to obtain absolute effects

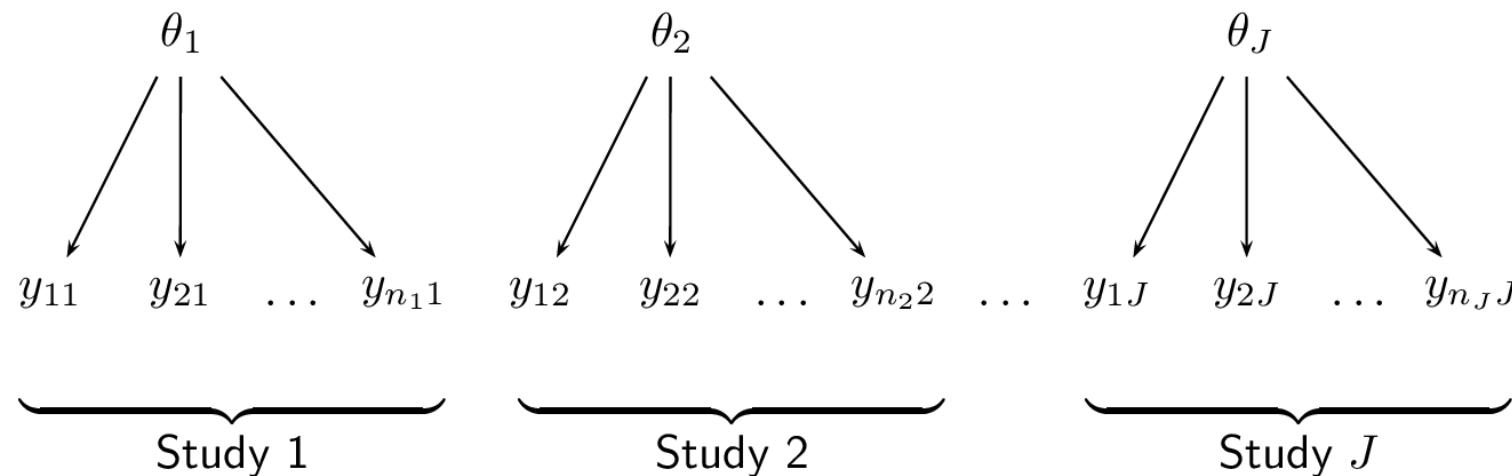
- $\text{mean}_1 = \text{mean}_0 + \text{mean difference}$
- $\text{log-odds}_1 = \text{log-odds}_0 + \text{log-odds ratio}$
- $\text{logit}(p_1) = \text{logit}(p_0) + \text{logOR}$
 - Transform from log-odds, μ , to probability, p , using inverse logit (sometimes called "expit"):

$$p = \frac{e^\mu}{1 + e^\mu} \iff \mu = \text{logit}(p) = \log\left(\frac{p}{1 - p}\right)$$

- $\text{log-hazard}_1 = \text{log-hazard}_0 + \text{logHR}$

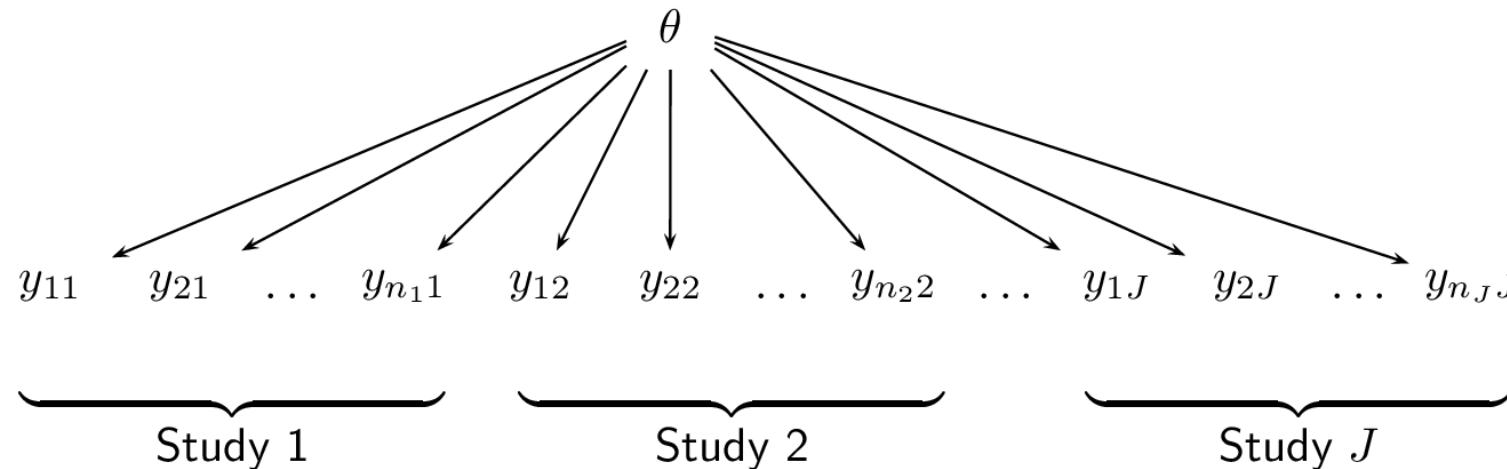
NB: we need to fully propagate the uncertainty in the relative to the absolute effect! (Being Bayesian makes life **much** easier!)

No pooling



- Assumes that relative effects from different studies are **independent**
 - Not much use for economic modelling, unless only a single study is relevant

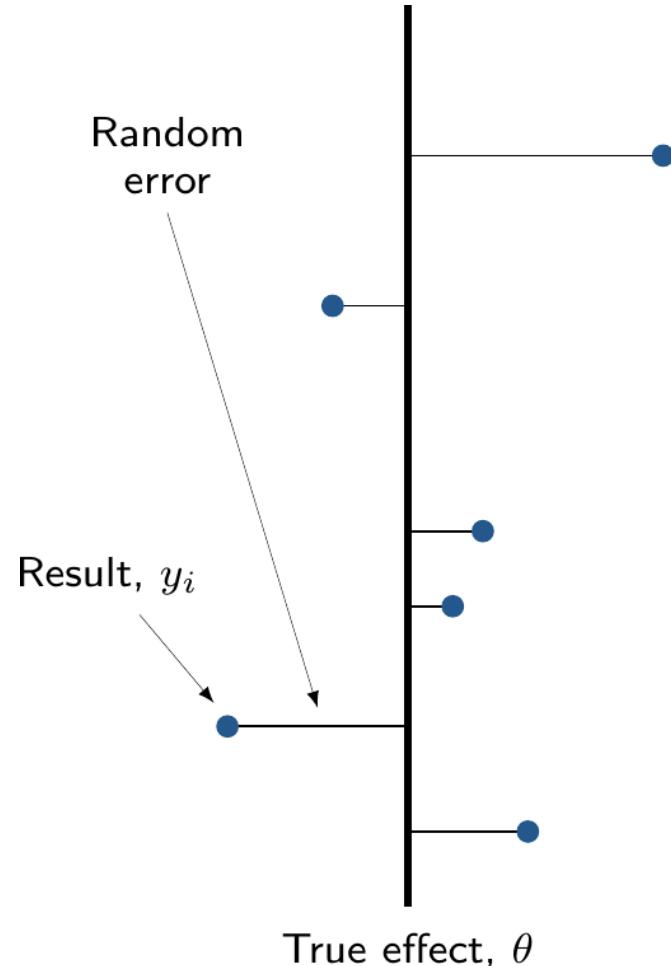
Complete pooling



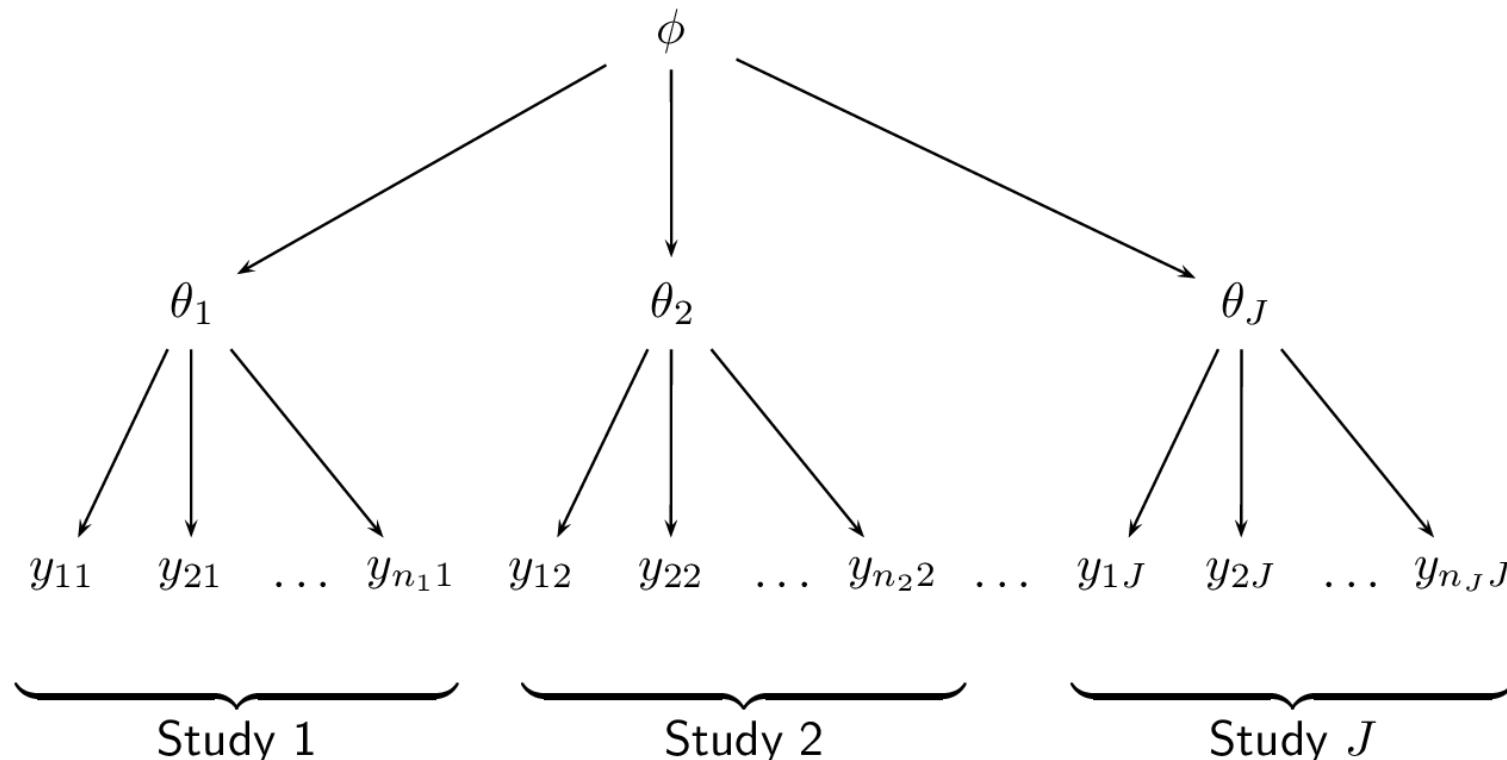
- Assumes that all studies (and all data points!) are **exchangeable**
 - They are "similar" – all come from the same data generating process
 - DGP depends on a single parameter θ

Complete pooling

- Statistical homogeneity
- Often called "**Fixed Effect**" model
- ... or "**Common Effect**" model



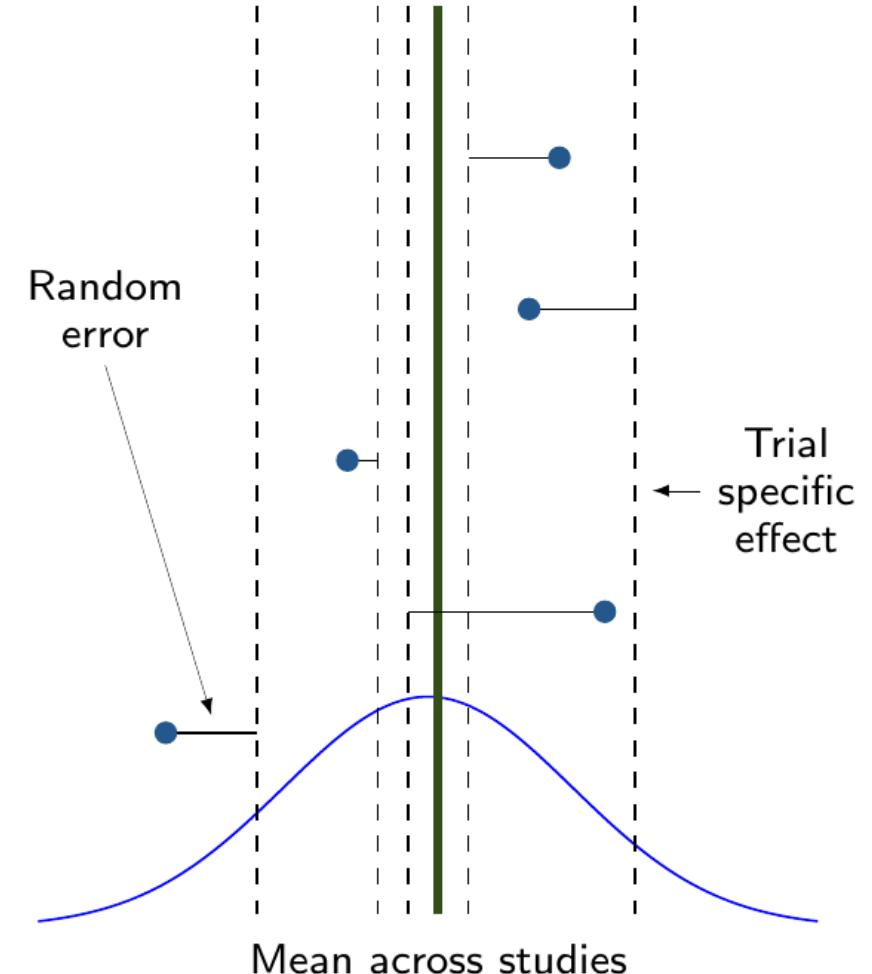
Partial pooling



- **Compromise**
 - Data inform a study-specific parameter
 - But all study-specific parameters are also exchangeable ("similar") and inform an "overall" parameter ϕ

Partial pooling

- Statistical heterogeneity
- Often called "**Random Effect**" model
- Estimate mean effect across studies and between study sd



- Partial pooling implies **hierarchical priors**

$$p(\theta_1, \theta_2, \dots, \theta_J, \phi) = p(\phi) \prod_{j=1}^J p(\theta_j \mid \phi)$$

- In a hierarchical model, the parameters are correlated \Rightarrow lower number of effective parameters!
- "Shrinkage"
 - ➔ See [this slide](#)

Example (Magnesium trial)

- Meta-analysis of 16 trials of magnesium vs placebo for patients with myocardial infarction

Trial	Trial name	Year	Magnesium		Placebo	
			Deaths (r_2)	Total (n_2)	Deaths (r_1)	Total (n_1)
1	Morton	1984	1	40	2	36
2	Rasmussen	1986	9	135	23	135
3	Smith	1986	2	200	7	200
4	Abraham	1987	1	48	1	46
5	Felstedt	1988	10	150	8	148
6	Shechter	1989	1	59	9	56
7	Ceremuzynski	1989	1	25	3	23
8	Bertschat	1989	0	22	1	21
9	Singh	1990	6	76	11	75
10	Pereira	1990	1	27	7	27
11	Shechter 1	1991	2	89	12	80
12	Golf	1991	5	23	13	33
13	Thorgersen	1991	4	130	8	122
14	LIMIT-2	1992	90	1159	118	1157
15	Shechter 2	1995	4	107	17	108
16	ISIS-4	1995	2216	29011	2103	29039

- For study i in arm k (for $k = 1$: placebo or $k = 2$: magnesium)

$$r_{1i} \sim \text{Binomial}(\pi_{1i}, n_{1i}) \quad r_{2i} \sim \text{Binomial}(\pi_{2i}, n_{2i})$$

$$\text{logit}(\pi_{ki}) = \alpha_i \quad \text{if } k = 1$$

$$\text{logit}(\pi_{ki}) = \alpha_i + \delta_i \quad \text{if } k = 2$$

Common effect model: $\delta_i = d$

Random effects model: $\delta_i \sim \text{Normal}(d, \sigma^2)$

- Independent priors given to α_i, d, σ

- For study i in arm k (for $k = 1$: placebo or $k = 2$: magnesium)

$$r_{1i} \sim \text{Binomial}(\pi_{1i}, n_{1i}) \quad r_{2i} \sim \text{Binomial}(\pi_{2i}, n_{2i})$$

$$\text{logit}(\pi_{ki}) = \alpha_i \quad \text{if } k = 1$$

$$\text{logit}(\pi_{ki}) = \alpha_i + \delta_i \quad \text{if } k = 2$$

Common effect model: $\delta_i = d$

Random effects model: $\delta_i \sim \text{Normal}(d, \sigma^2)$

- Independent priors given to α_i, d, σ
- Can easily obtain summary for OR

$$\text{OR} = e^d$$

- For random effects model, can summarise in 2 ways:
 - Mean across all study effects, d
 - Prediction for a "new" population, exchangeable with included study populations:

$$\delta_{pred} \sim \text{Normal}(d, \sigma^2)$$

More on this in Higgins & Spiegelhalter, 2002

No pooling model

```
model{  
  for(i in 1:ns) {  
    # Sampling distributions  
    r1[i] ~ dbin(pi1[i], n1[i])  
    r2[i] ~ dbin(pi2[i], n2[i])  
    # Model  
    logit(pi1[i]) <- alpha[i]  
    logit(pi2[i]) <- alpha[i] + delta[i]  
    # Baselines  
    alpha[i] ~ dnorm(0.0,1.0E-5)  
    # Study-specific treatment effects  
    delta[i] ~ dnorm(0.0,1.0E-5)  
  }  
}
```

Random effects model

```
model{  
  for(i in 1:ns) {  
    # Sampling distributions  
    r1[i] ~ dbin(pi1[i], n1[i])  
    r2[i] ~ dbin(pi2[i], n2[i])  
    # Model  
    logit(pi1[i]) <- alpha[i]  
    logit(pi2[i]) <- alpha[i] + delta[i]  
    # Baselines  
    alpha[i] ~ dnorm(0.0,1.0E-5)  
    # Random Effects  
    delta[i] ~ dnorm(d, prec)  
  }  
  
  # Priors  
  d ~ dnorm(0.0,1.0E-6)  
  sd ~ dunif(0,5)  
  prec <- pow(sd,-2)  
  
  # Summaries  
  OR <- exp(d)  
  delta.pred ~ dnorm(d, prec)  
}
```

Common effects model

```
model{  
  for(i in 1:ns) {  
    # Sampling distributions  
    r1[i] ~ dbin(pi1[i], n1[i])  
    r2[i] ~ dbin(pi2[i], n2[i])  
    # Model  
    logit(pi1[i]) <- alpha[i]  
    logit(pi2[i]) <- alpha[i] + delta[i]  
    # Baselines  
    alpha[i] ~ dnorm(0.0,1.0E-5)  
    # "Degenerate" Random Effects  
    delta[i] ~ dnorm(d, prec)  
  }  
  
  # Priors  
  d ~ dnorm(0.0,1.0E-6)  
  sd <- 0      # No variance for the random effect (so not really random...)  
  prec <- pow(sd,-2)  
  
  # Summaries  
  OR <- exp(d)  
  delta.pred ~ dnorm(d, prec)  
}
```

Common effects model

```
model{  
  for(i in 1:ns) {  
    # Sampling distributions  
    r1[i] ~ dbin(pi1[i], n1[i])  
    r2[i] ~ dbin(pi2[i], n2[i])  
    # Model  
    logit(pi1[i]) <- alpha[i]  
    logit(pi2[i]) <- alpha[i] + d  
    # Baselines  
    alpha[i] ~ dnorm(0.0,1.0E-5)  
  }  
  
  # Priors  
  d ~ dnorm(0.0,1.0E-6)  
  
  # Summaries  
  OR <- exp(d)  
}
```

- Results very different for random/common effect models
- Fixed effects model gives higher **Deviance Information Criterion** (DIC) even allowing for model complexity

$$DIC = p_D + \overline{D(\theta)} = D(\bar{\theta}) + 2p_D$$

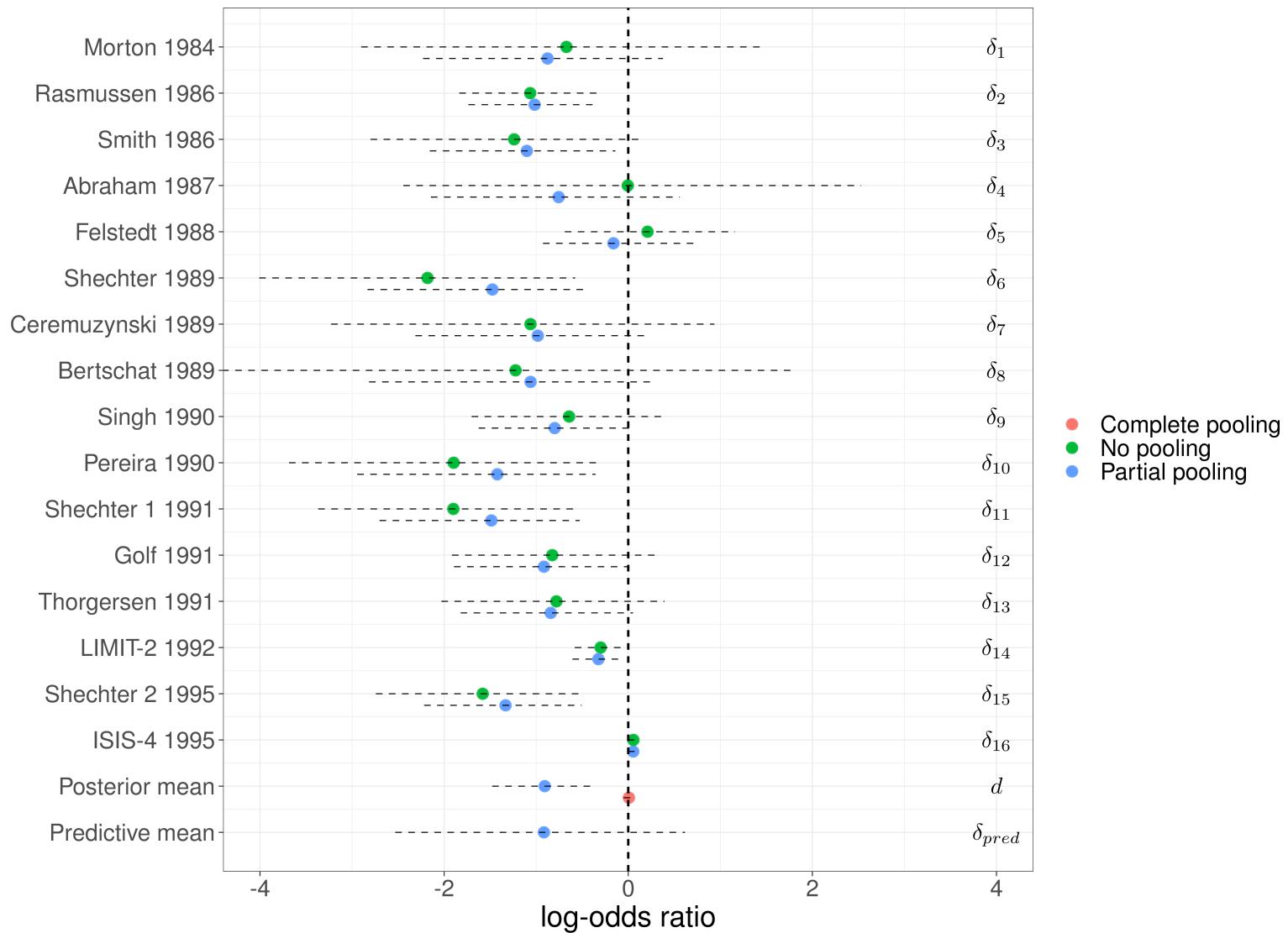
$\overline{D(\theta)}$ = model mean deviance (see [Lecture 5](#))

$D(\bar{\theta})$ = model deviance at the average value of parameters

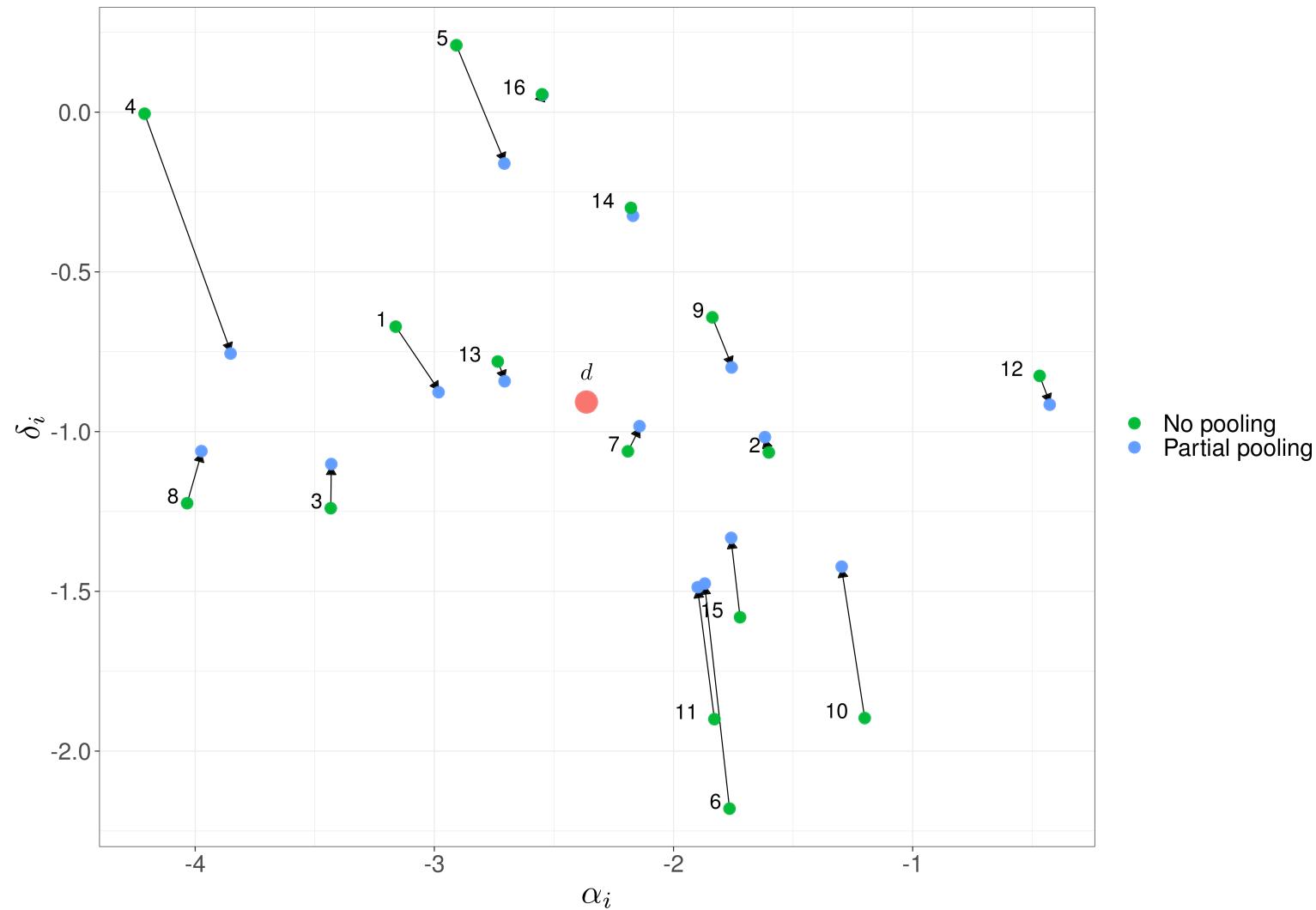
$p_D = \overline{D(\theta)} - D(\bar{\theta})$ = "effective" number of model parameters (more on this later)

Model	OR (95% interval)	Between study sd (σ)	Posterior mean deviance (\bar{D})	p_D	DIC
Random effects	0.42 (0.23 - 0.66)	0.72 (0.36 - 1.33)	147.0	24.2	171.2
Common effects	1.01 (0.95 - 1.07)		195.1	16.6	211.8

Results



Posterior means for the model parameters

 Back

Referred to as p_D

- Common Effect Model
 - p_D = number of parameters
- Random Effects Model
 - p_D depends on between-study standard deviation – it is a measure of **shrinkage**
 - For sd close to 0, $\delta_i = d$; 1 parameter (as in common effect model – complete pooling)
 - For very large sd, $\delta_i = \delta_i$; parameter for each study (as in independent effects model – no pooling)

How many effective parameters in the magnesium example?

Fixed effects model ($sd = 0$): $[p_D = 16.6]$

α_i , 16 studies

d , common treatment effects

Random effects model ($sd > 0$) $[p_D = 24.2]$

α_i , 16 studies

δ_i , exchangeable (partial pooling)

Independent effects model ($sd \rightarrow \infty$) $[p_D = 28.5]$

α_i , 16 studies

δ_i , 16 independent treatment effects

17 parameters



Up to 32 parameters

Effective number of parameters

The computation of p_D is based on an **approximation**.

- The original paper defining the DIC ( [Spiegelhalter et al, 2002](#)) uses the definition

$$p_D = \overline{D(\theta)} - D(\bar{\theta})$$

- That is not the only possible approximation.  [Gelman et al \(2004\)](#) suggested

$$p_D = \frac{\text{Var}[D(\theta)]}{2}$$

which is *invariant to parameterisations* of the model (and sometimes referred to as p_V – see slide 7 of  [this](#))

Effective number of parameters

The computation of p_D is based on an **approximation**.

- The original paper defining the DIC ( [Spiegelhalter et al, 2002](#)) uses the definition

$$p_D = \overline{D(\theta)} - D(\bar{\theta})$$

- That is not the only possible approximation.  [Gelman et al \(2004\)](#) suggested

$$p_D = \frac{\text{Var}[D(\theta)]}{2}$$

which is *invariant to parameterisations* of the model (and sometimes referred to as p_V – see slide 7 of  [this](#))

- R2openBUGS takes advantage of the fact that OpenBUGS can compute both versions and gives the first one, by default
- R2jags (and JAGS) only use the second version

Effective number of parameters

The computation of p_D is based on an **approximation**.

- The original paper defining the DIC ( [Spiegelhalter et al, 2002](#)) uses the definition

$$p_D = \overline{D(\theta)} - D(\bar{\theta})$$

- That is not the only possible approximation.  [Gelman et al \(2004\)](#) suggested

$$p_D = \frac{\text{Var}[D(\theta)]}{2}$$

which is *invariant to parameterisations* of the model (and sometimes referred to as p_V – see slide 7 of  [this](#))

- R2openBUGS takes advantage of the fact that OpenBUGS can compute both versions and gives the first one, by default
- R2jags (and JAGS) only use the second version

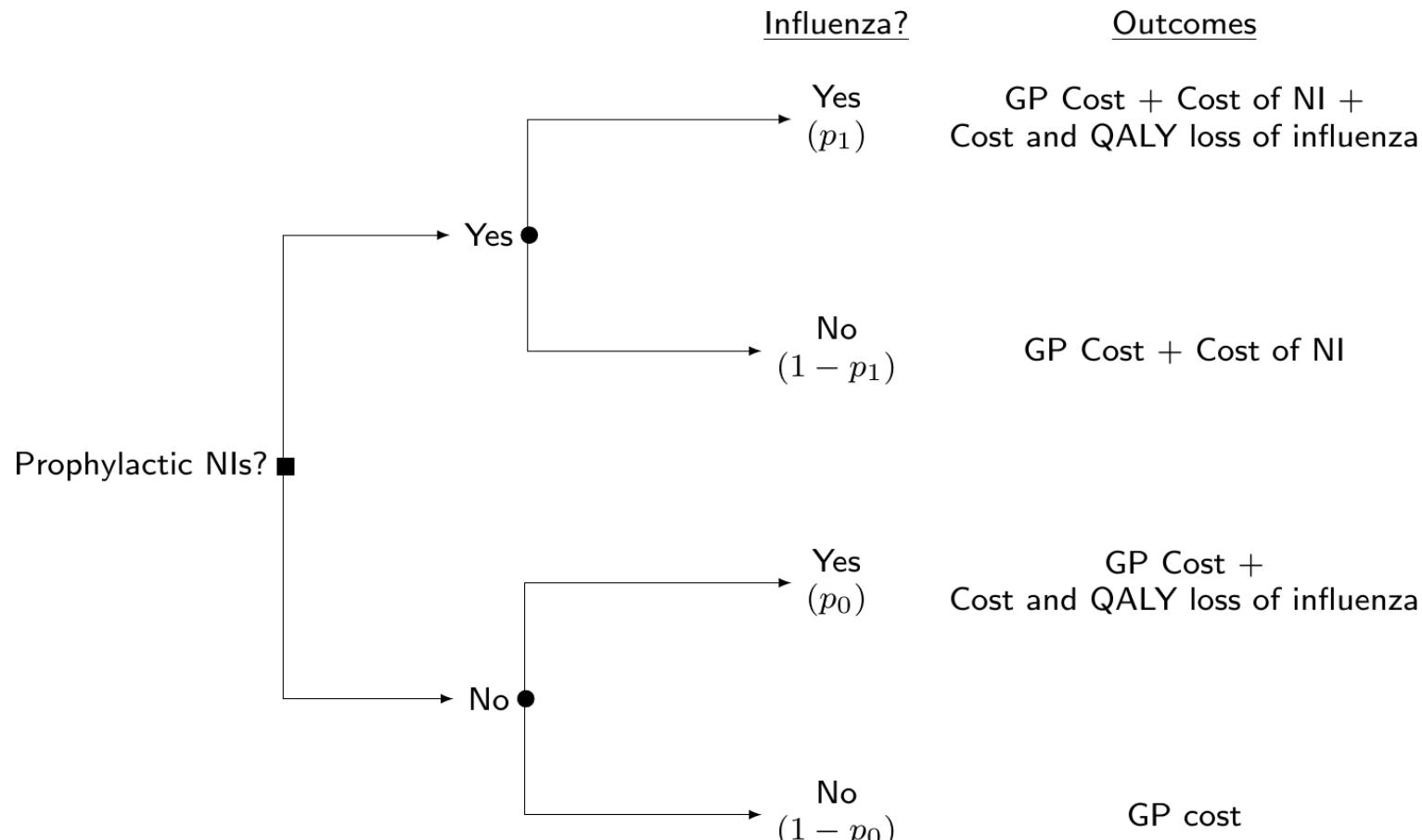
In the magnesium example, the two software give **completely** different answers

- In R2jags the distribution of $D(\theta)$ is affected by a few *outliers* (simulations with **huge** values). This makes the estimate of its variance unstable and very large, thus generating ridiculous values for p_D
- Conversely, if you run the model with R2openBUGS, p_D is computed using the first formula and returns sensible values
- See also  [here](#) for a technical discussion of the implementation in R2openBUGS

Prior for the between study SD

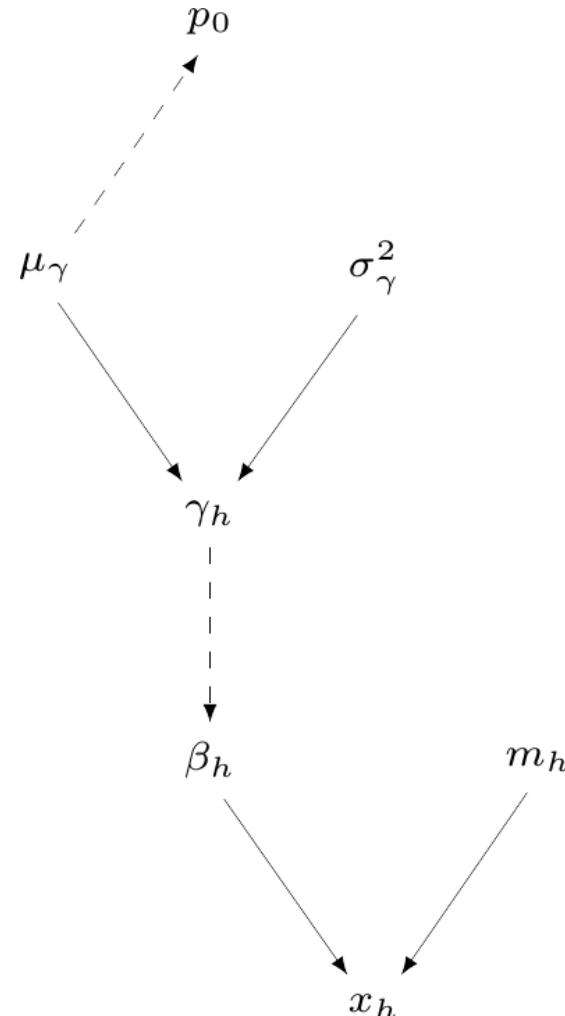
- Priors for between study sd , σ , are tricky!
 - Must be positive, but $\sigma = 0$ is feasible... most priors biased away from zero
- Typically limited data to estimate σ
 - Prior weight on high values can be problematic
 - Informative prior based on previous meta-analyses can be helpful ([Turner et al 2015](#))
- Results can be sensitive to choice of prior
 - Uniform prior for σ over a realistic range often a good choice
 - But check posterior not constrained by prior...
 - Consider sensitivity analysis

Prophylactic Neuraminidase Inhibitors (NIIs) for Influenza



- Absolute effect on reference (no NIIs) informed by synthesis of H studies
- Relative effect for NI vs No NI informed by synthesis of RCTs

Decision-analytic model (influenza)



Module 1: Influenza incidence

- H studies reporting number of patients who get influenza (x_h) in the sample (m_h)
- β_h = population probability of influenza from the h -th study:

$$\text{logit}(\beta_h) = \gamma_h \sim \text{Normal}(\mu_\gamma, \sigma_\gamma)$$

- $\mu_\gamma \sim \text{Normal}(0, v)$ = pooled averaged probability of infection (on logit scale!)

$$\Rightarrow p_0 = \frac{\exp(\mu_\gamma)}{1 + \exp(\mu_\gamma)}$$

(or equivalently $\text{logit}(p_0) = \mu_\gamma$)

Decision-analytic model (influenza)

Module 2: Prophylaxis effectiveness

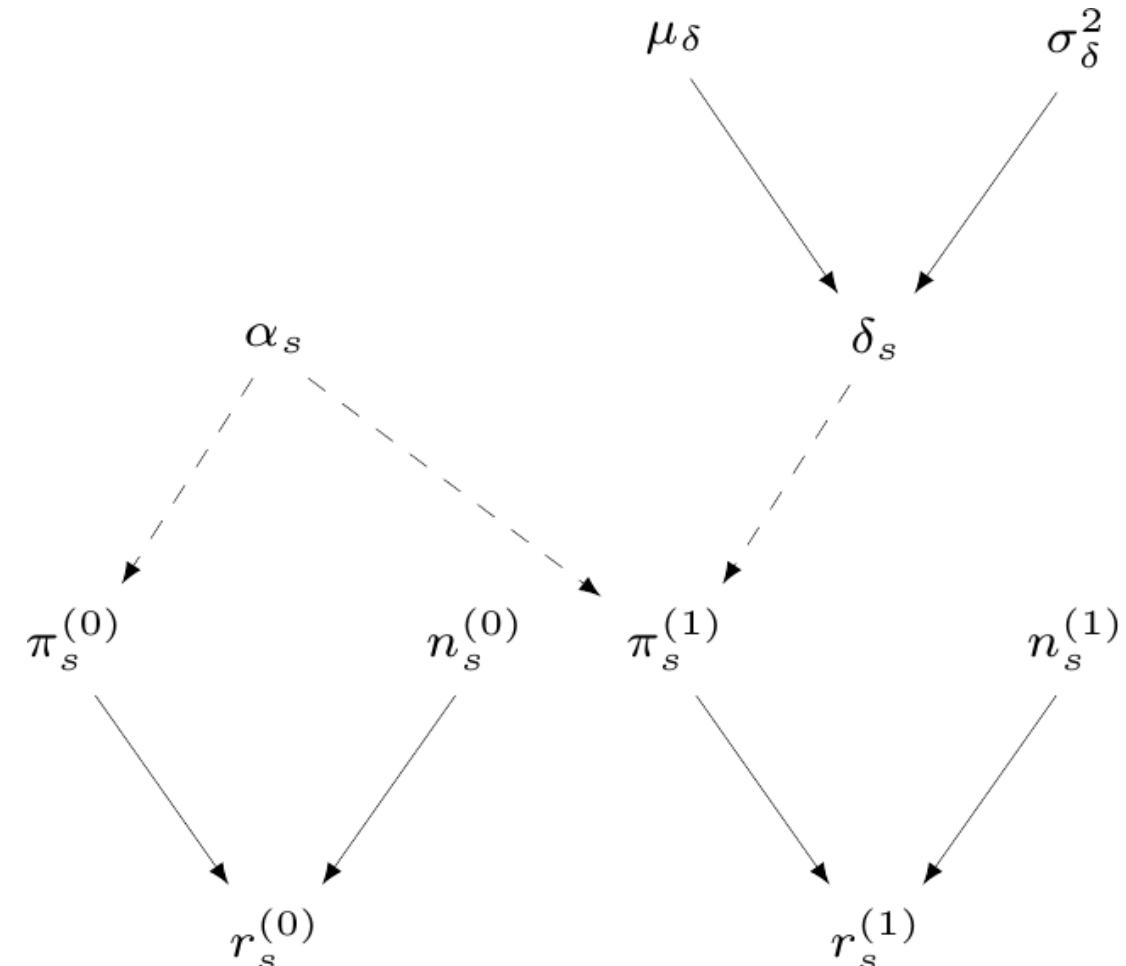
- S studies reporting number of infected patients $r_s^{(t)}$ in a sample made of $n_s^{(t)}$ subjects
- $\pi_s^{(t)} = \text{study- and treatment-specific chance of contracting influenza}$

$$\text{logit}(\pi_s^{(0)}) = \alpha_s \sim \text{Normal}(0, 10)$$

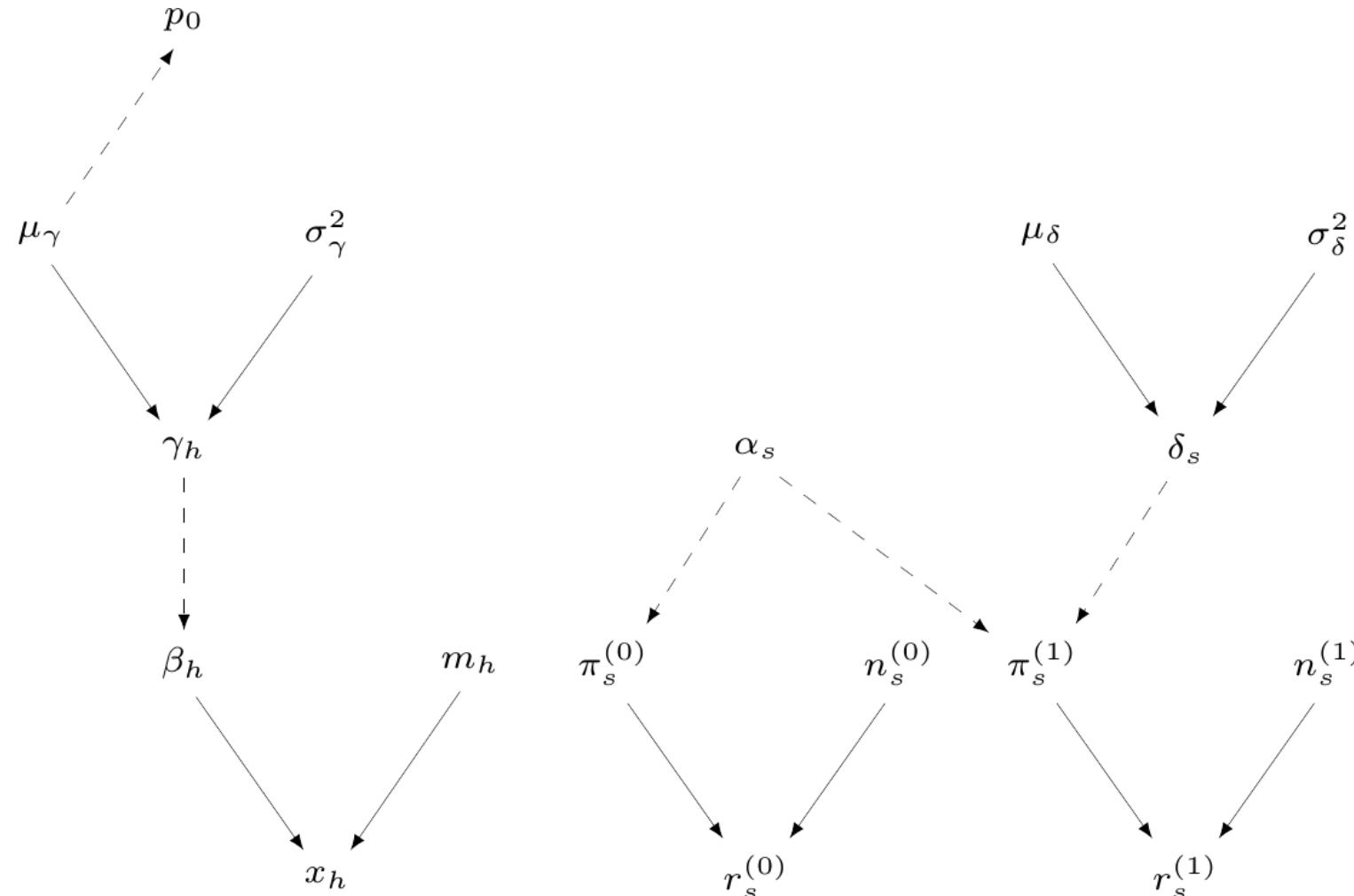
$$\text{logit}(\pi_s^{(1)}) = \alpha_s + \delta_s$$

$\delta_s \sim \text{Normal}(\mu_\delta, \sigma_\delta) = \text{study-specific treatment effect}$

- $\mu_\delta \sim \text{Normal}(0, v) = \text{pooled log-odds ratio of influenza given treatment}$



Decision-analytic model (influenza)



Can combine modules 1 and 2
 $\text{logit}(p_1) = \text{logit}(p_0) + \mu_\delta$

Decision-analytic model (influenza)

BUGS code

```
model {  
  # Evidence synthesis on incidence of influenza in the  
  # "healthy" adults population (t=0)  
  for (h in 1:H) {  
    x[h] ~ dbin(beta[h], m[h])  
    logit(beta[h]) <- gamma[h]  
    gamma[h] ~ dnorm(mu.gamma, tau.gamma)  
  }  
}
```

Decision-analytic model (influenza)

BUGS code

```
model {  
  # Evidence synthesis on incidence of influenza in the  
  # "healthy" adults population (t=0)  
  for (h in 1:H) {  
    x[h] ~ dbin(beta[h], m[h])  
    logit(beta[h]) <- gamma[h]  
    gamma[h] ~ dnorm(mu.gamma, tau.gamma)  
  }  
  
  # Evidence synthesis for effectiveness of NIs (t=1 vs t=0)  
  for (s in 1:S) {  
    r0[s] ~ dbin(pi0[s], n0[s])  
    r1[s] ~ dbin(pi1[s], n1[s])  
    logit(pi0[s]) <- alpha[s]  
    logit(pi1[s]) <- alpha[s]+delta[s]  
    delta[s] ~ dnorm(mu.delta, tau.delta)  
    alpha[s] ~ dnorm(0, 0.0001)  
  }  
}
```

Decision-analytic model (influenza)

BUGS code

```
# Prior distributions
mu.delta ~ dnorm(0,0.0001);          mu.gamma ~ dnorm(0,0.0001)
sigma.delta ~ dunif(0,10);             tau.delta <- pow(sigma.delta,-2)
sigma.gamma ~ dunif(0,10);             tau.gamma <- pow(sigma.gamma,-2)

# Costs of influenza
c.inf ~ dnorm(mu.inf,tau.inf)

# Length of time to recovery when infected by influenza
l ~ dlnorm(mu.l,tau.l)

# Odds Ratio of influenza under treatment with NIs
rho <- exp(mu.delta)

# Estimated probability of influenza in "healthy adults" for t=0
logit(p0) <- mu.gamma

# Estimated probability of influenza in "healthy adults" for t=1
logit(p1) <- mu.gamma + mu.delta
}
```

Run model code

Model fitting

Economic analysis

Correlations

Decision analysis

- See practical!

[Run model code](#)[Model fitting](#)[Economic analysis](#)[Correlations](#)[Decision analysis](#)

```
> # ... Run the script to fit the evidence synthesis model, something like:  
> # model=jags(...)  
>  
> # Then prints the output for selected parameters  
> print(model,intervals=c(.025,.975),digits=3)
```

Inference for Bugs model at "/tmp/Rtmpf4Vjp1/file6876f5b49f718", fit using jags,
2 chains, each with 10000 iterations (first 5000 discarded), n.thin = 10
n.sims = 1000 iterations saved

	mu.vect	sd.vect	2.5%	97.5%	Rhat	n.eff
c.inf	16.874	2.282	12.569	21.473	1.001	1000
l	8.170	1.452	5.676	11.376	1.000	1000
p1	0.060	0.023	0.026	0.107	1.005	1000
p2	0.014	0.009	0.005	0.031	1.001	1000
rho	0.217	0.087	0.109	0.410	1.007	480
deviance	102.202	5.751	92.523	114.642	1.001	1000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)

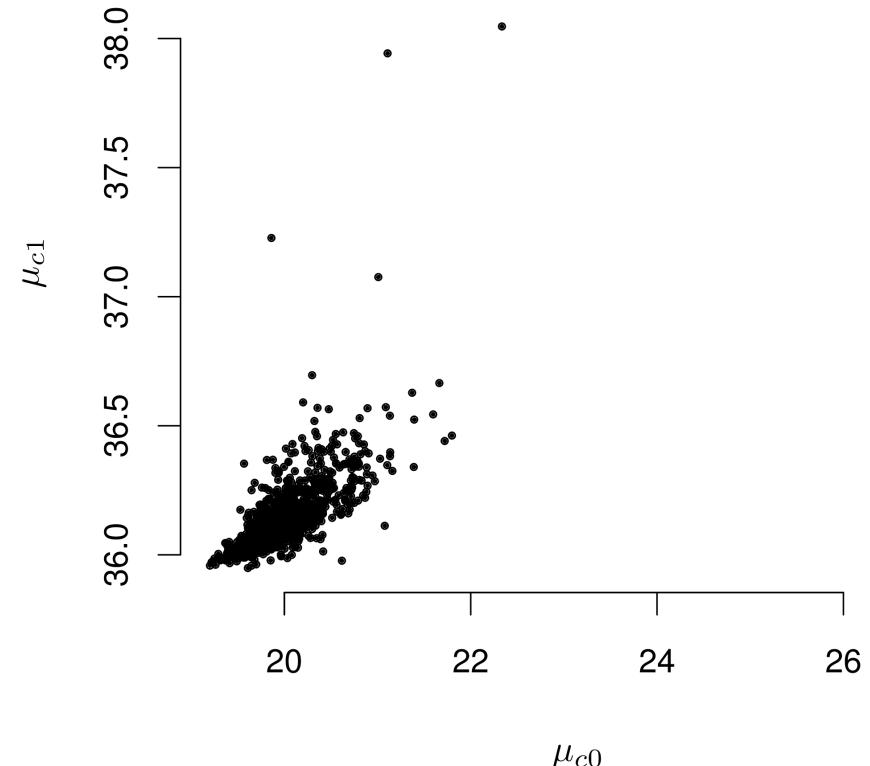
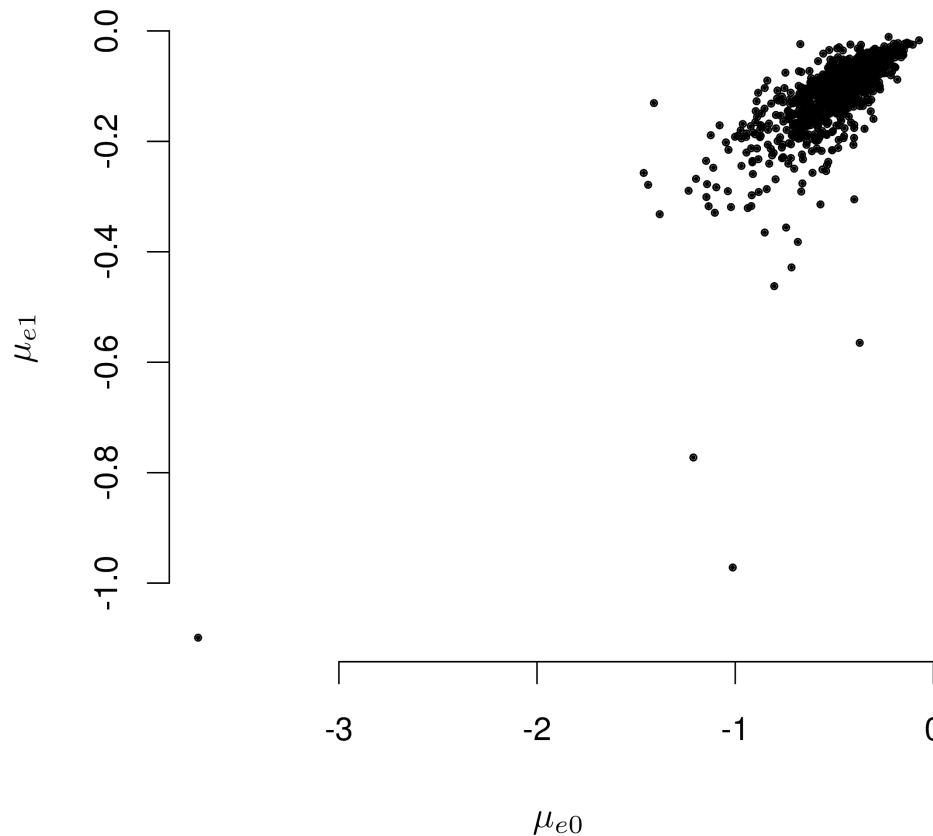
pD = 16.5 and DIC = 118.7

DIC is an estimate of expected predictive error (lower deviance is better).

[Run model code](#)[Model fitting](#)[Economic analysis](#)[Correlations](#)[Decision analysis](#)

```
> # Combines output from MCMC model to define population average effectiveness & costs
> attach.jags(model)
> # Cost of treatment
> c <- e <- matrix(NA,n.sims,T)
> c[,1] <- (1-p1)*(c_gp) + p1*(c_gp+c_inf)
> c[,2] <- (1-p2)*(c_gp+c_ni) + p2*(c_gp+c_ni+c_inf)
>
> # Measure of effectiveness
> e[,1] <- -l*p1
> e[,2] <- -l*p2
>
> # These are mu_e and mu_c and can be used to do the decision analysis!
```

- l = proxy for QALY loss due to influenza (length of time to recovery)
- **NB:** GP cost incurred under all options, and so could be omitted from model...

[Run model code](#)[Model fitting](#)[Economic analysis](#)[Correlations](#)[Decision analysis](#)

[Run model code](#)[Model fitting](#)[Economic analysis](#)[Correlations](#)[Decision analysis](#)

```
> library(BCEA)
> trt.labels=c("status quo",
+               "prophylaxis with NI
+ )
> m <- bcea(e,c,ref=2,
+             interventions=trt.labels
+             Kmax=10000
+ )
> summary(m)
```

NB: k (wtp) is defined in the interval [0 - 10000]

Cost-effectiveness analysis summary

Reference intervention: prophylaxis with NI
Comparator intervention: status quo

Optimal decision: choose status quo for $k < 60$ and prophylaxis with NI if $k > 60$

Analysis for willingness to pay parameter $k = 10000$

Expected utility

status quo -4947.4

prophylaxis with NI -1168.3

EIB CEAC ICER

prophylaxis with NI vs status quo 3779.1 0.999 42.52

Optimal intervention (max expected utility) for $k = 10000$: prophylaxis with NI

EVPI 1891.5