Bayesian Statistics for Genetics

Lecture 8: Meta-analysis

July, 2020

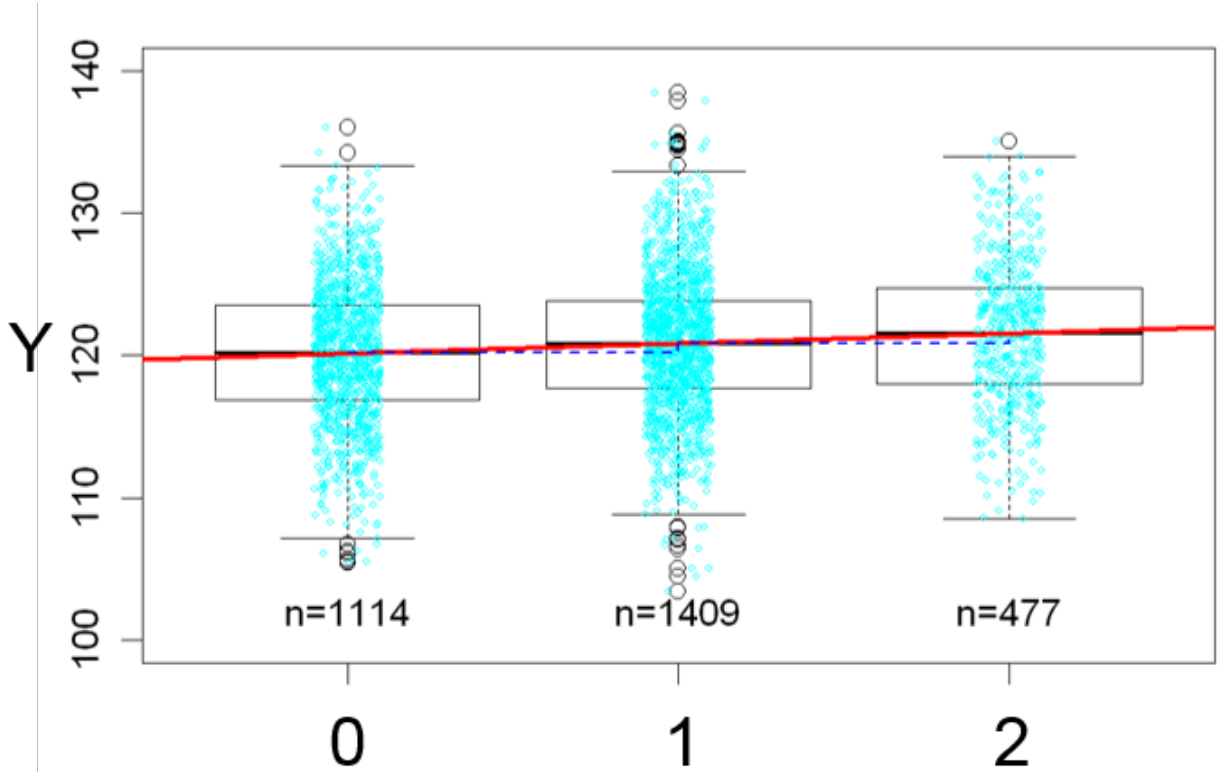
Overview

The ability to combine information from multiple sources is a strength of Bayesian statistics;

- Use of prior information + study data
- Combining multiple studies' data, in a *meta-analysis*

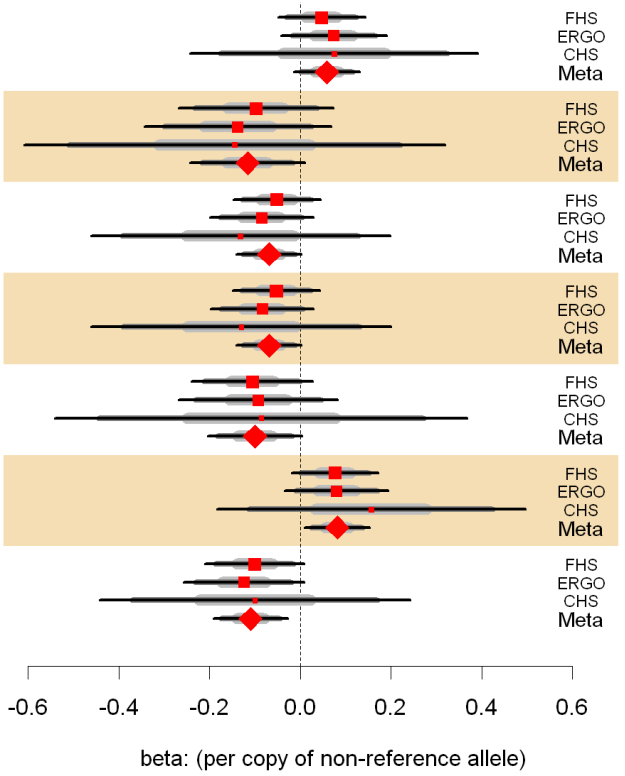
Meta-analysis was briefly introduced in Session 7 (GLMs) – here we give a more general approach, and look at *mixed models*, which are also natural in Bayesian approaches.

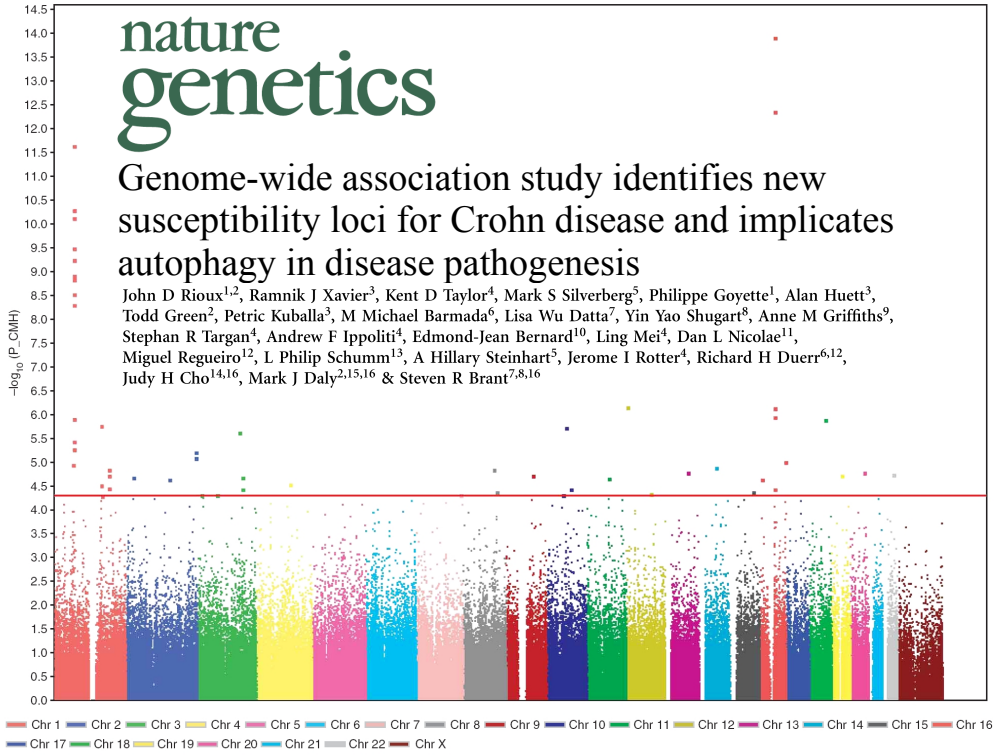
Overview



- Test association of disease (Y) with genotype ($X = 0/1/2$) – is there a signal? (If so, learn new biology)
- Tiny effects, so combine multiple studies – meta-analysis

Overview





- Test association of disease (Y) with genotype ($X = 0/1/2$) – is there a signal? (If so, learn new biology)
- Tiny effects, so combine multiple studies – meta-analysis

Meta-analysis: default approaches

In medium-sized studies or larger, can approximate the data from study i well, as

$$\widehat{\beta}_i \sim N(\beta_i, \sigma_i^2),$$

where each study is big enough that uncertainty about σ_i is negligible.

How about across studies? One very simple model assumes *homogeneity*, i.e.

$$\beta_i = \beta_0$$

and, with a flat prior on common parameter β_0, Session 4's results tell us that

$$\begin{aligned} \widehat{\beta}_F = \mathbb{E}[\,\beta_0|\text{data}\,] &= \sum_{i=1}^{k} \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \widehat{\beta}_i, \\ \text{Var}[\,\widehat{\beta}_F\,] = \text{Var}[\,\beta_0|\text{data}\,] &= \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}. \end{aligned}$$

Meta-analysis: default approaches

This is known as the *fixed-effect* or *common-effect* approach, and $\widehat{\beta}_F$ is the *inverse-variance weighted* or *precision-weighted* estimate

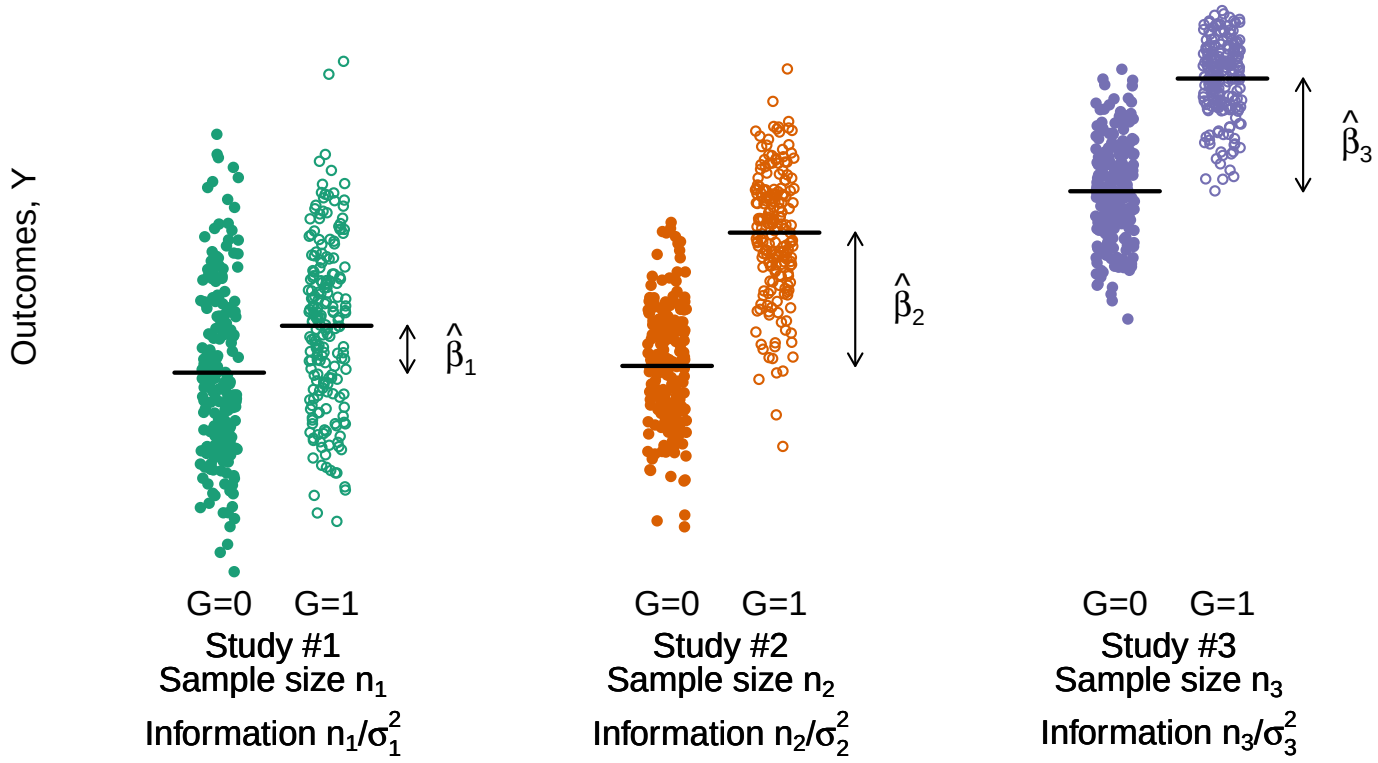
- Under true homogeneity, just as efficient as pooling all the data and adjusting for study (Lin & Zeng, 2010)
- Under true homogeneity, Uniformly Most Powerful Unbiased (i.e. best) estimate of β_0

But

- True homogeneity is not generally plausible – perhaps in lab replicates, perhaps if all $\beta_i = 0$
- Not learning about *heterogeneity* – which may be important in practice

Meta-analysis: under heterogeneity

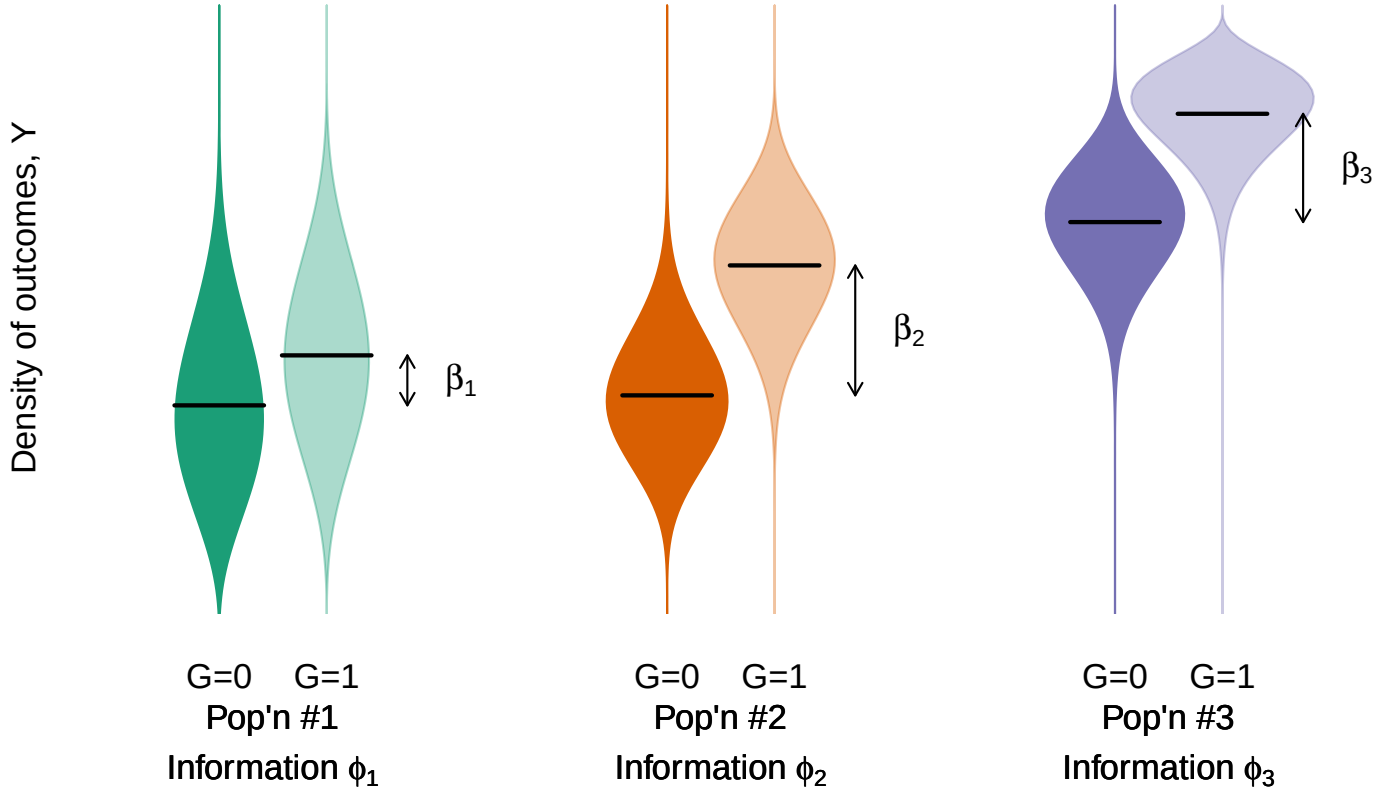
To help think about $\beta_i \neq \beta_0$, consider *data* from three studies;



Each $n_i = 200$ here. We assume all σ_i^2 known ...can relax this.

Meta-analysis: under heterogeneity

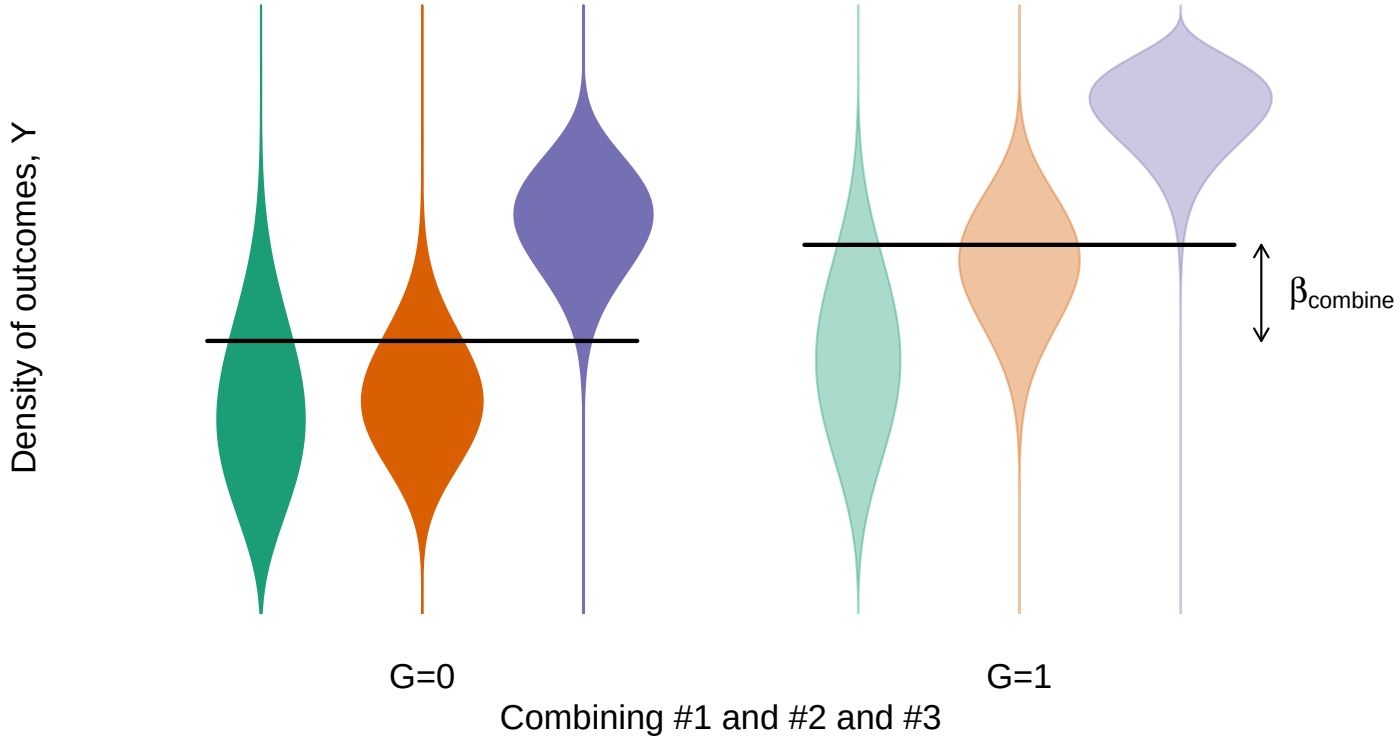
Parameters those 3 studies are estimating;



Differences in means (β_i) *and* information per observation (ϕ_i)

Meta-analysis: under heterogeneity

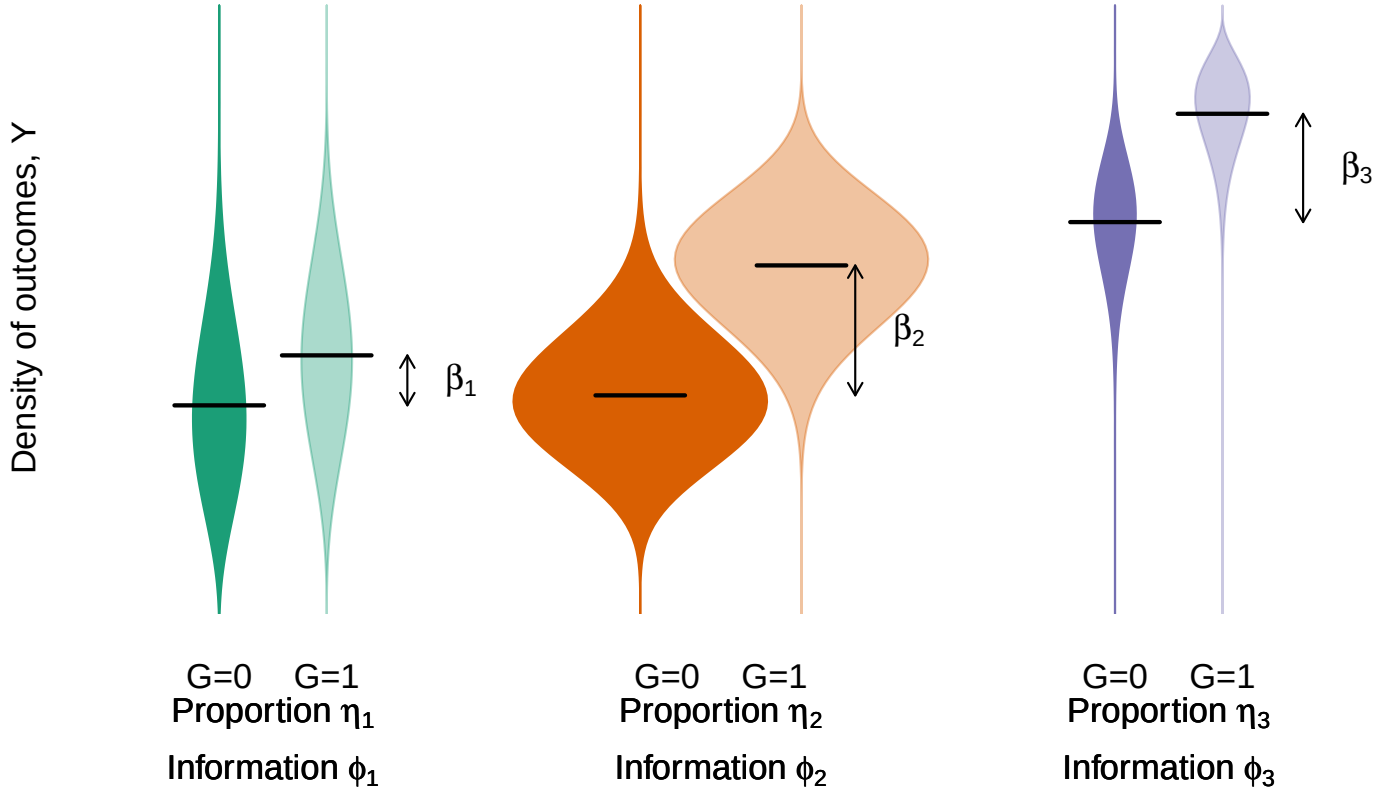
One overall population we might learn about;



Mean difference, with each sub-population represented equally.

Meta-analysis: under heterogeneity

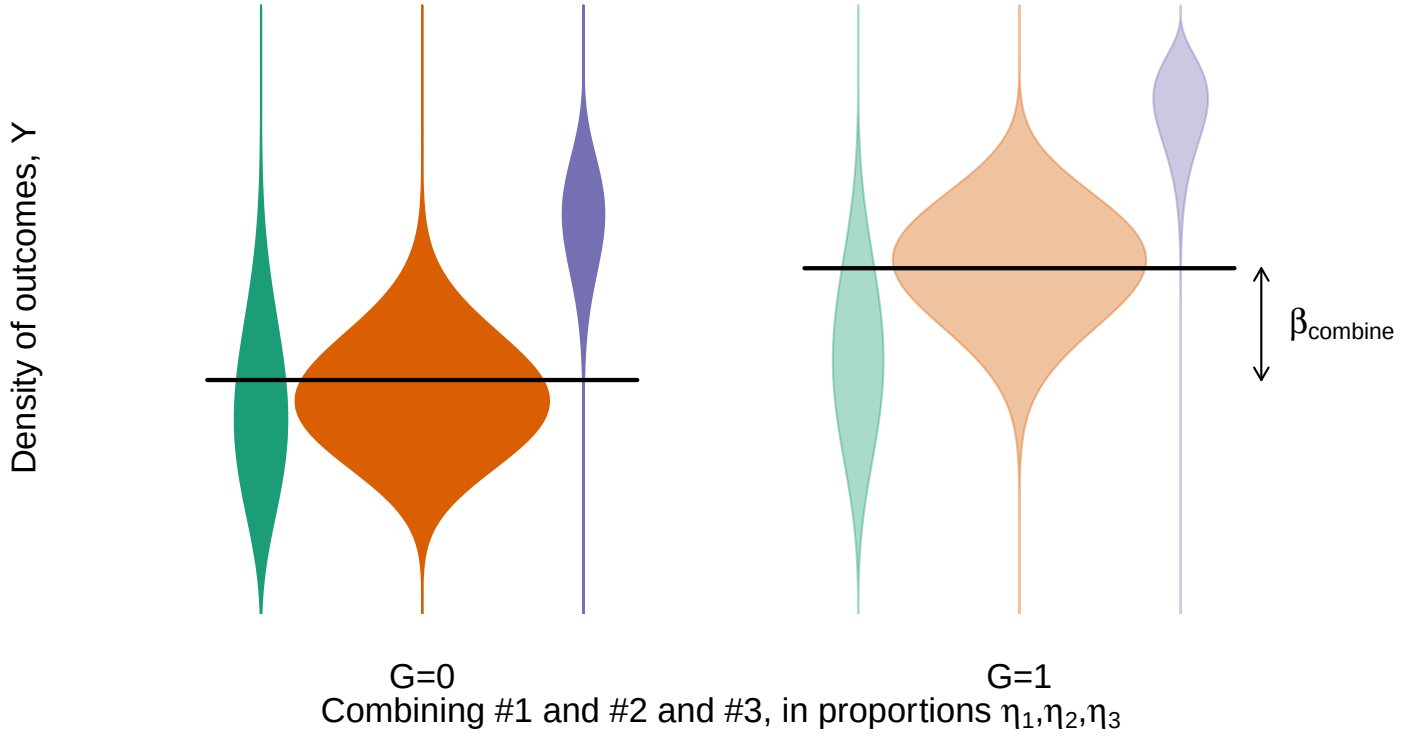
Another overall population we might learn about;



Weights here are 2/7/1, not 1/1/1 as before.

Meta-analysis: under heterogeneity

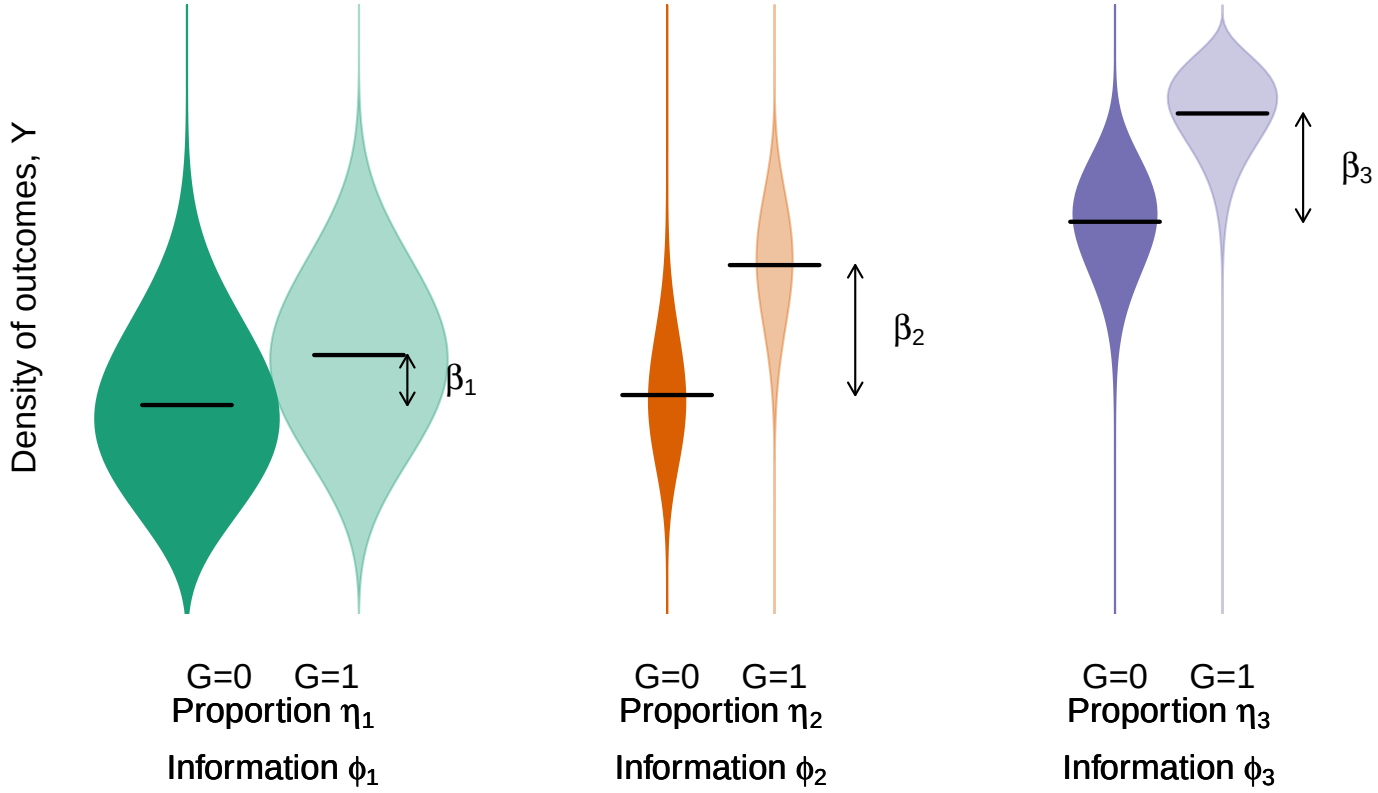
Another overall population we might learn about;



Still an average effect, but closer to β_2 than before.

Meta-analysis: under heterogeneity

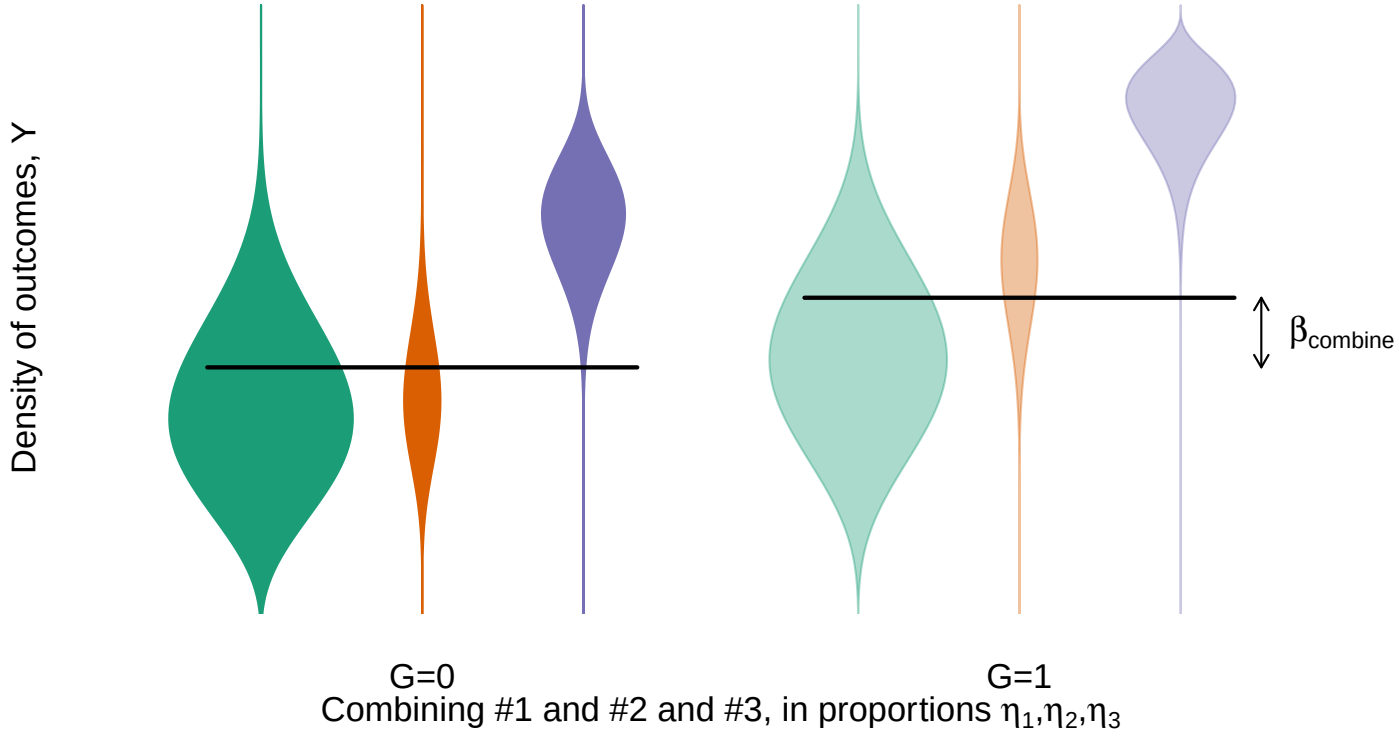
And another; (obviously, there are unlimited possibilities)



Weights here are 7/1/2.

Meta-analysis: under heterogeneity

And another; (obviously, there are unlimited possibilities)



Weights here are 7/1/2 – smaller average effect, closer to β_1

Meta-analysis: under heterogeneity

With a flat prior, among all the weighted averages which has smallest posterior variance? The answer may look familiar;

$$\beta_F = \sum_{i=1}^{k} \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \beta_i$$

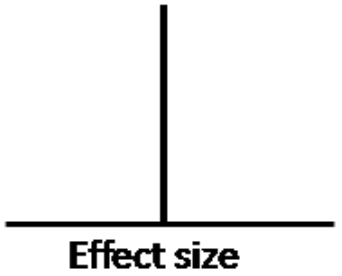
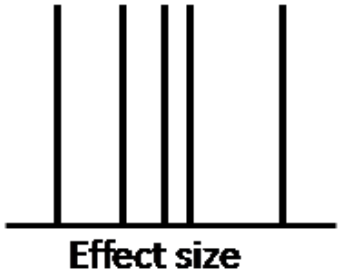
Its posterior mean and variance are *the same* as we saw for common effect β_0

- Known as the *fixed-effect***S** approach (note the plural) – it assumes one fixed effect for each study, we estimate an average
- ... the average the data tells us most about

> A single estimator can have more than one valid justification. If this applies to your estimator, state why *you* are using it.

Meta-analysis: under heterogeneity

Those justifications once again;

Name:	Common effect	Fixed effect**s**
Assumptions:		
	All $\beta_i = \beta_0$	β_i unrestricted
Plausible?	**Seldom**	**Often**
$\widehat{\beta}_F$ estimates:	Single β_0	An average, β_F
Valid estimate?	Yes	Yes
Var[$\widehat{\beta}_F$] valid?	≈Yes*	≈Yes*

- When testing, only care if **all** $\beta_i = 0$, when common-effect=fixed-effects
- This area is surprisingly controversial...

* Having negligible error in σ_i matters, but small-sample "fixes" are available

Meta-analysis: under heterogeneity

Q. Can I use $\widehat{\beta}_F$ under heterogeneity? A. It depends who you ask (!)

... it is no longer an estimate of any parameter, nor can its standard error or associated confidence interval be found

Whitehead & Whitehead, SiM

The assumption should thus be viewed as a potentially useful approximation

Greenland & Rothman, Modern Epi, pg 270

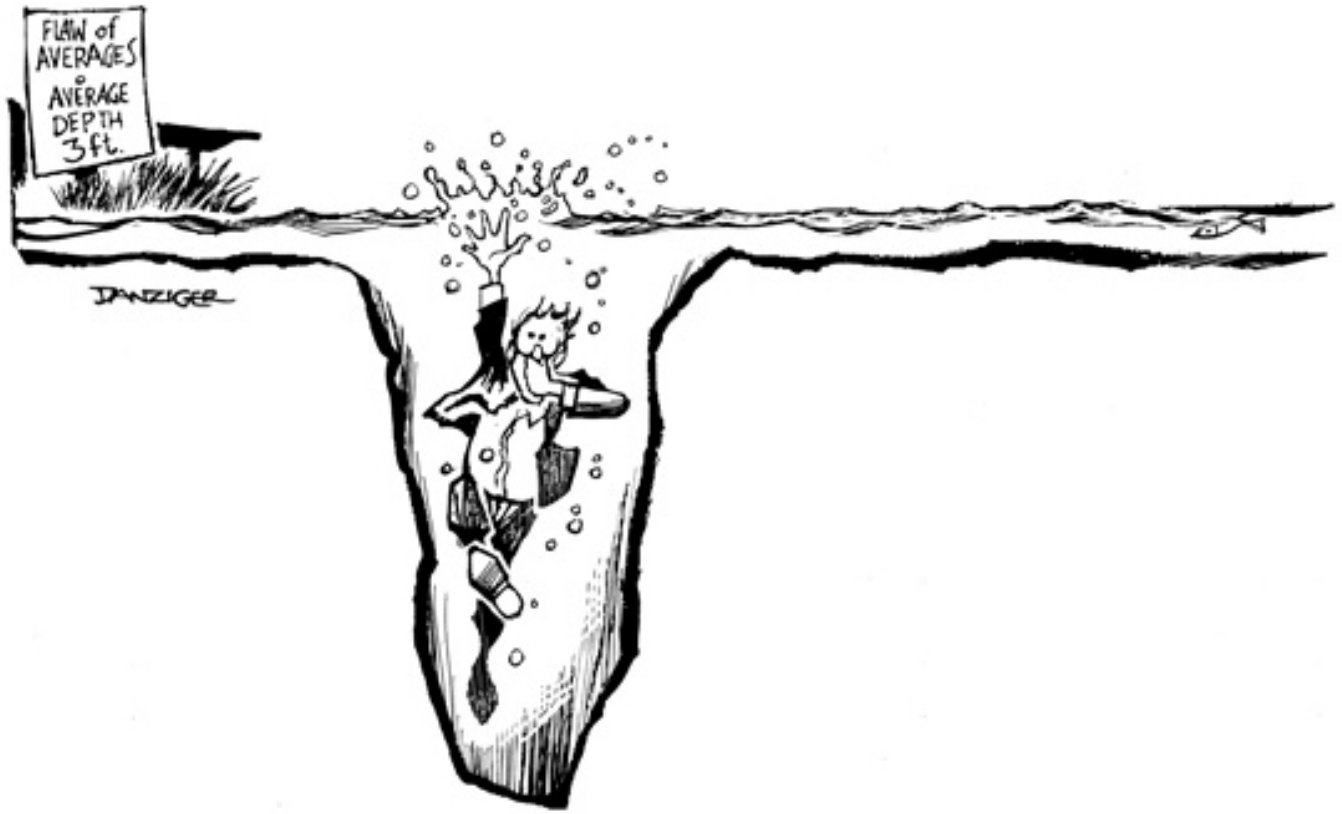
... it does **not**, *however, implicitly assume that the true effect of treatment is the same in each trial*

Peto *et al*, e.g. Lancet, 1998

Default advice makes users reluctant to report $\widehat{\beta}_F$ alone under heterogeneity.

Meta-analysis: under heterogeneity

Letting an average (e.g. β_F) tell the whole story is the '*flaw* of averages';



- Average effect β_F answers one question
- This does not mean other questions aren't interesting!

Meta-analysis: under heterogeneity

An obvious measure of 'dispersion', i.e. spread;

$$\frac{1}{k}\sum_{i=1}^{k}(\beta_i - \beta_F)^2.$$

But we actually learn more about a weighted average of deviations around β_F;

$$\zeta^2 = \frac{1}{\sum_{i=1}^{k}\eta_i\phi_i}\sum_{i=1}^{k}\eta_i\phi_i(\beta_i - \beta_F)^2.$$

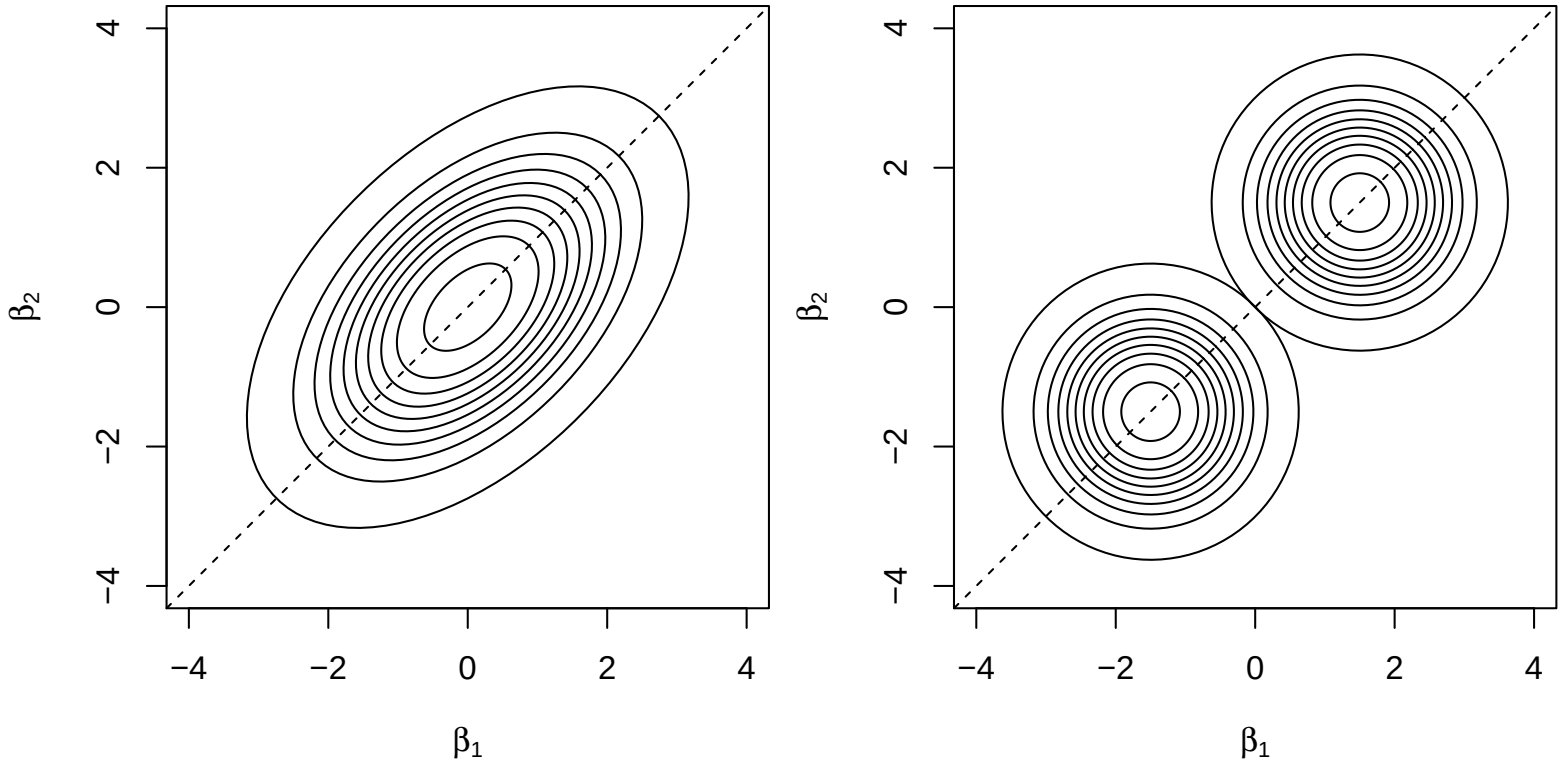
An empirical estimate of this quantity can be written

$$\tilde{\zeta}^2 = \frac{\sum_{i=1}^{k}\sigma_i^{-2}(\beta_i - \widehat{\beta}_F)^2 - (k-1)}{\sum_{i=1}^{k}\sigma_i^{-2}} = \frac{Q-(k-1)}{\sum_{i=1}^{k}\sigma_i^{-2}}$$

where Q is a.k.a. *Cochran's Q*, and $I^2 = 1-(k-1)/Q$ (trunacted at zero) are standard non-Bayesian statistics for testing homogeneity.

Meta-analysis: exchangeability

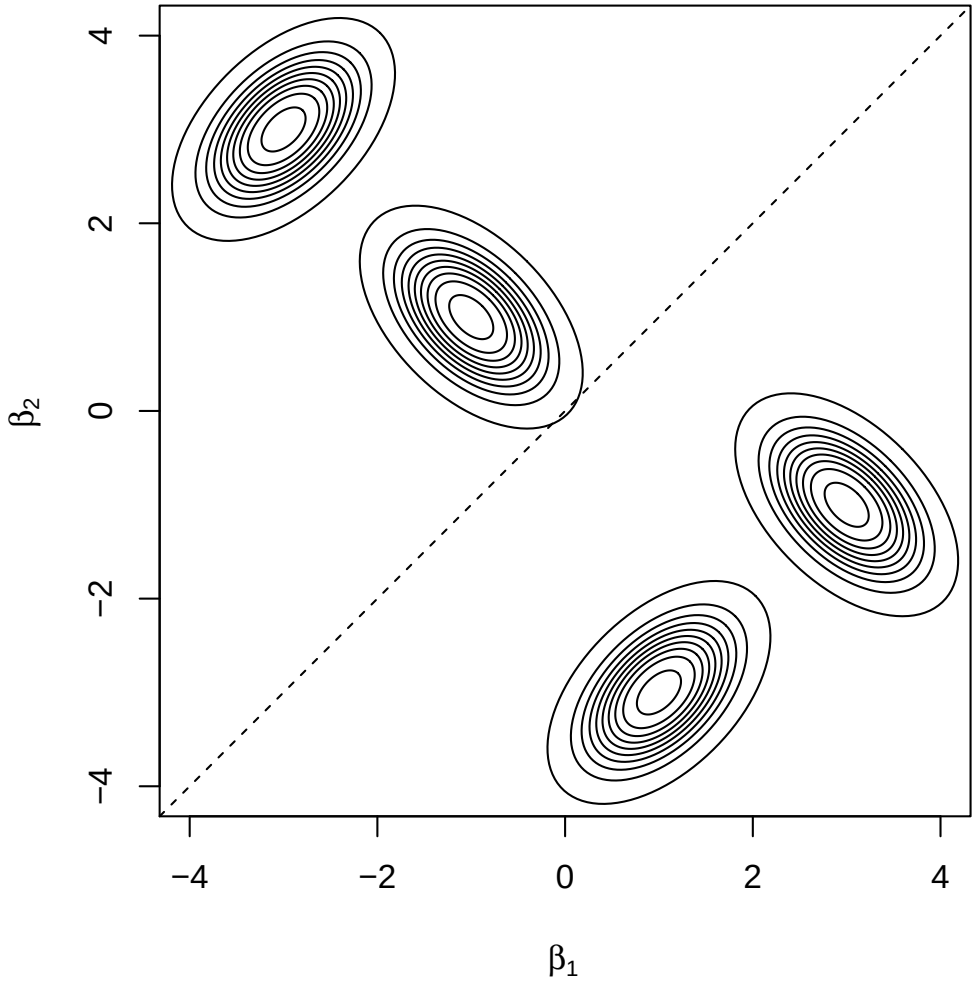
As we've seen, no prior really describes lack of knowledge But for multiple parameters, we can (easily) state that knowledge about them is symmetric;



The property $p(\beta_1, \beta_2) = p(\beta_2, \beta_1)$ is called *exchangeability*.

Meta-analysis: exchangeability

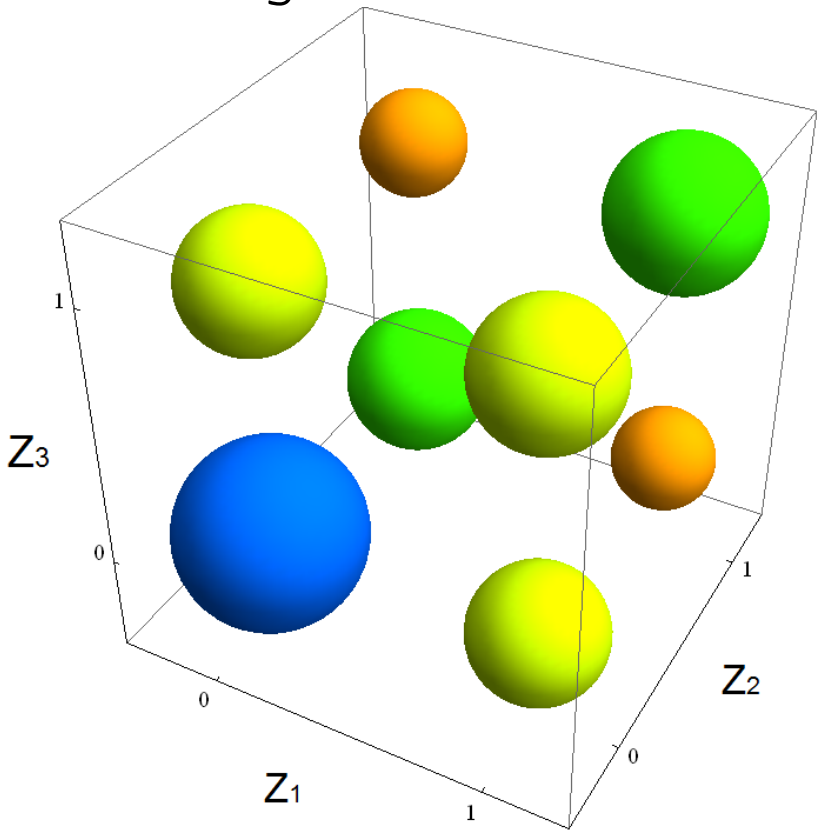
Exchangeability is a weaker statement than $p(\beta_1, \beta_2) = p(\beta_1)p(\beta_2)$, a.k.a. *independence* (see previous slide) and a stronger statement than having identical distributions (see right).



Meta-analysis: partial exchangeability (*)

An example with binary $\{z_1, z_2, z_3\}$; (colors indicate probabilities) Are z_1, z_2, z_3 identically distributed? Independent? Exchangeable?

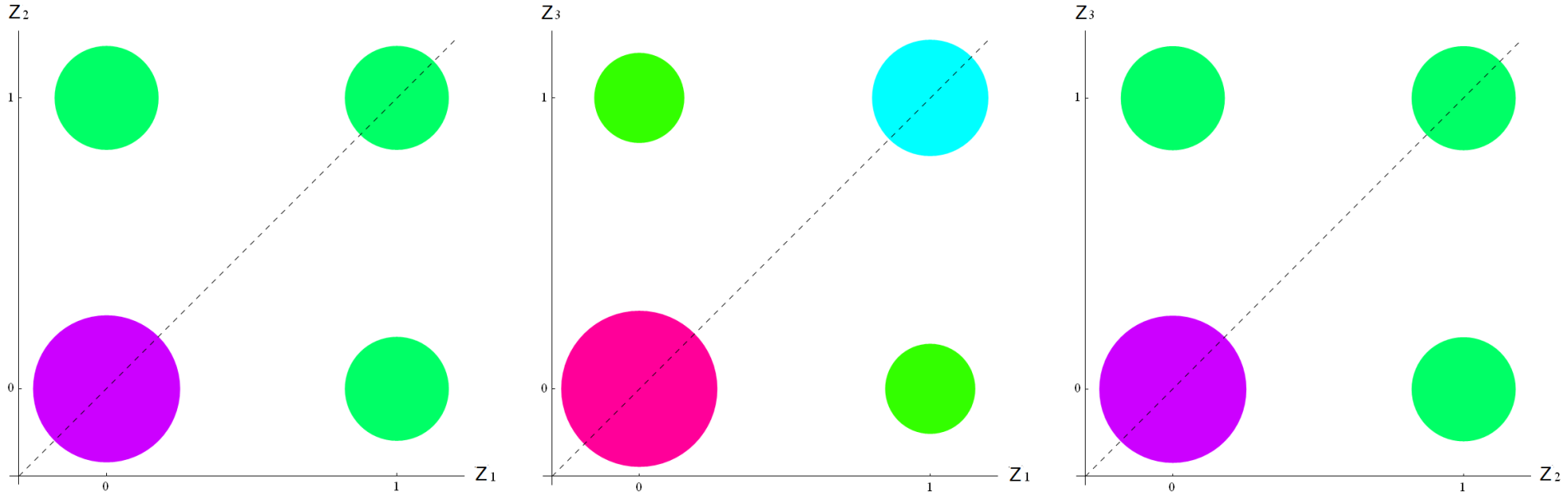
z_1	z_2	z_3	$\mathbb{P}[\boldsymbol{z}]$
0	0	0	6/20
0	0	1	2/20
0	1	0	3/20
0	1	1	1/20
1	0	0	2/20
1	0	1	2/20
1	1	0	1/20
1	1	1	3/20



Meta-analysis: partial exchangeability (*)

Full exchangeability does not hold, but any two variables are exchangeable; (colors indicate probabilities again)

Y_2 vs Y_1 | Y_3 vs Y_1 | Y_3 vs Y_2



The variables $\{Y_1, Y_2, Y_3\}$ are *2-exchangeable*; the concept can be generalized to n-exchangeability.

Meta-analysis: using exchangeability

In an exchangeable prior, there's no distinction between what we know about one β_i versus another. For example, a prior with exchangeable β_i, for meta-analysis;

$$\begin{aligned} \widehat{\beta}_i &\sim N(\beta_i, \sigma_i^2) \\ \beta_i &\overset{i.i.d.}{\sim} N(\mu, \tau^2) \end{aligned}$$

...for some μ, τ^2 – which may in turn have *hyperpriors*, describing uncertainty about the prior for the β_i.

This is a form of *hierarchical model* – more on these in later sessions.

Remarkably, it turns out that Bayesian hierarchical models and exchangeability are **equivalent** – this is *de Finetti's theorem*.
If your beliefs on the β_i have exchangeability, they **must** correspond to β_i being random draws from some mixing distribution.

Meta-analysis: using exchangeability

In this hierarchical model, the default not-so-Bayesian estimate for μ is *Der Simonian-Laird* (DSL);

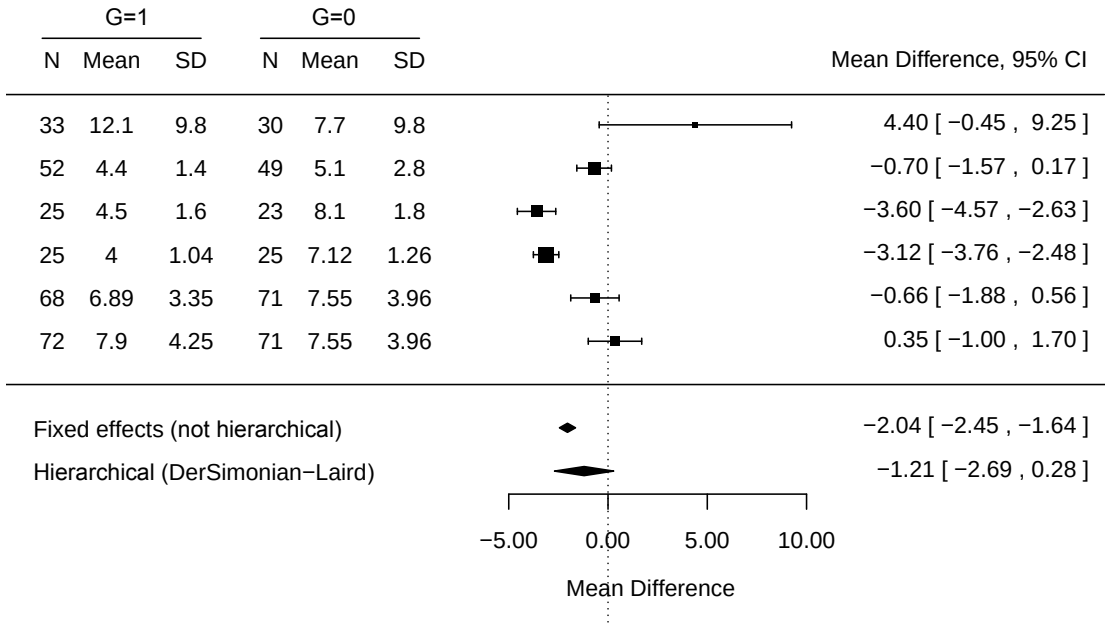
$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \hat{\tau}^2} \hat{\beta}_i}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \hat{\tau}^2}}, \qquad \text{with } \text{Var}[\,\hat{\beta}_F\,] = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \hat{\tau}^2}},$$

$$\text{and } \hat{\tau}^2 = \max\left(\frac{Q - (k-1)}{\sum \sigma_i^{-2} - \sum \sigma_i^{-4} / \sum \sigma_i^{-2}}, 0 \right)$$

- DSL uses a method of moments plug-in for τ^2, then fairly natural
- Gives $\hat{\beta}_F$ when Q (heterogeneity) is below-average compared to homogeneity
- Estimates a weighted average of the β_i – but where inverse-variance weights are 'moderated' by τ^2

Meta-analysis: example

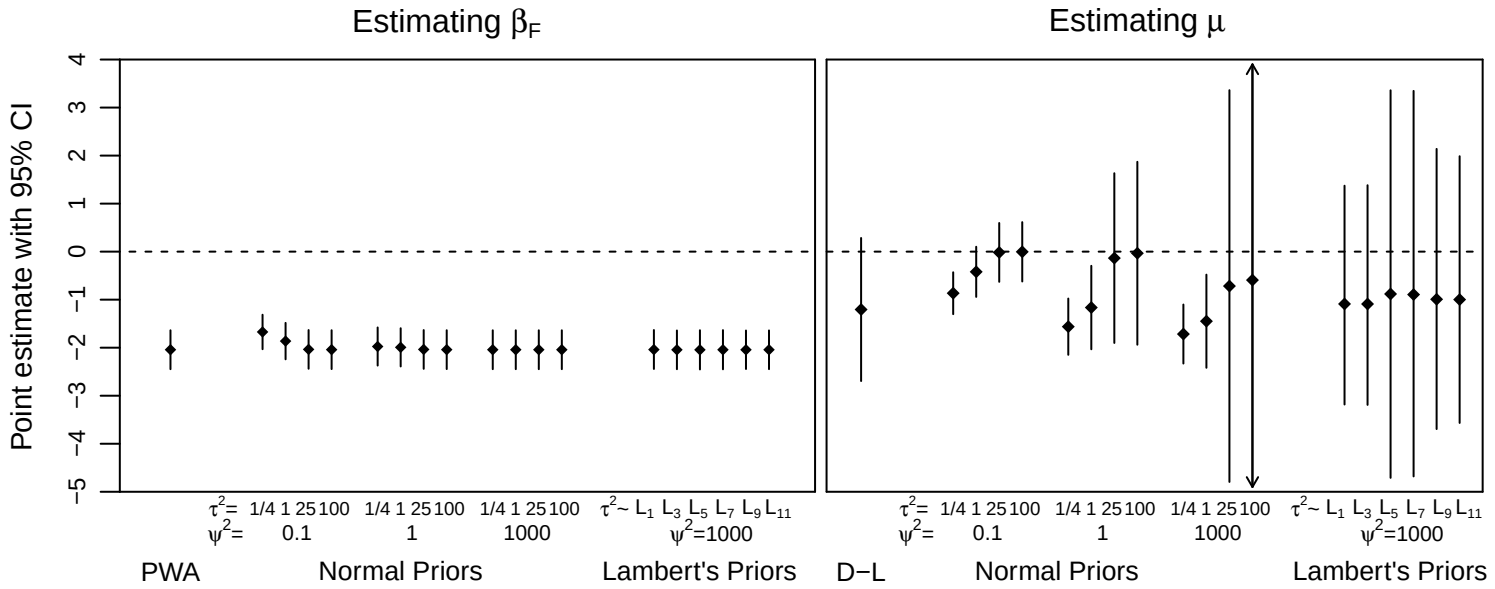
Typical meta-analysis of 5 association studies;



Using full Bayes, we can introduce priors on the hyper-parameters;

$$
\begin{aligned}
\widehat{\beta}_i &\sim N(\beta_i, \sigma_i^2) \\
\beta_i &\overset{i.i.d.}{\sim} N(\mu, \tau^2) \\
\mu &\sim N(0, \psi^2) \\
\tau^2 &\sim p(\tau^2)
\end{aligned}
$$

Meta-analysis: example

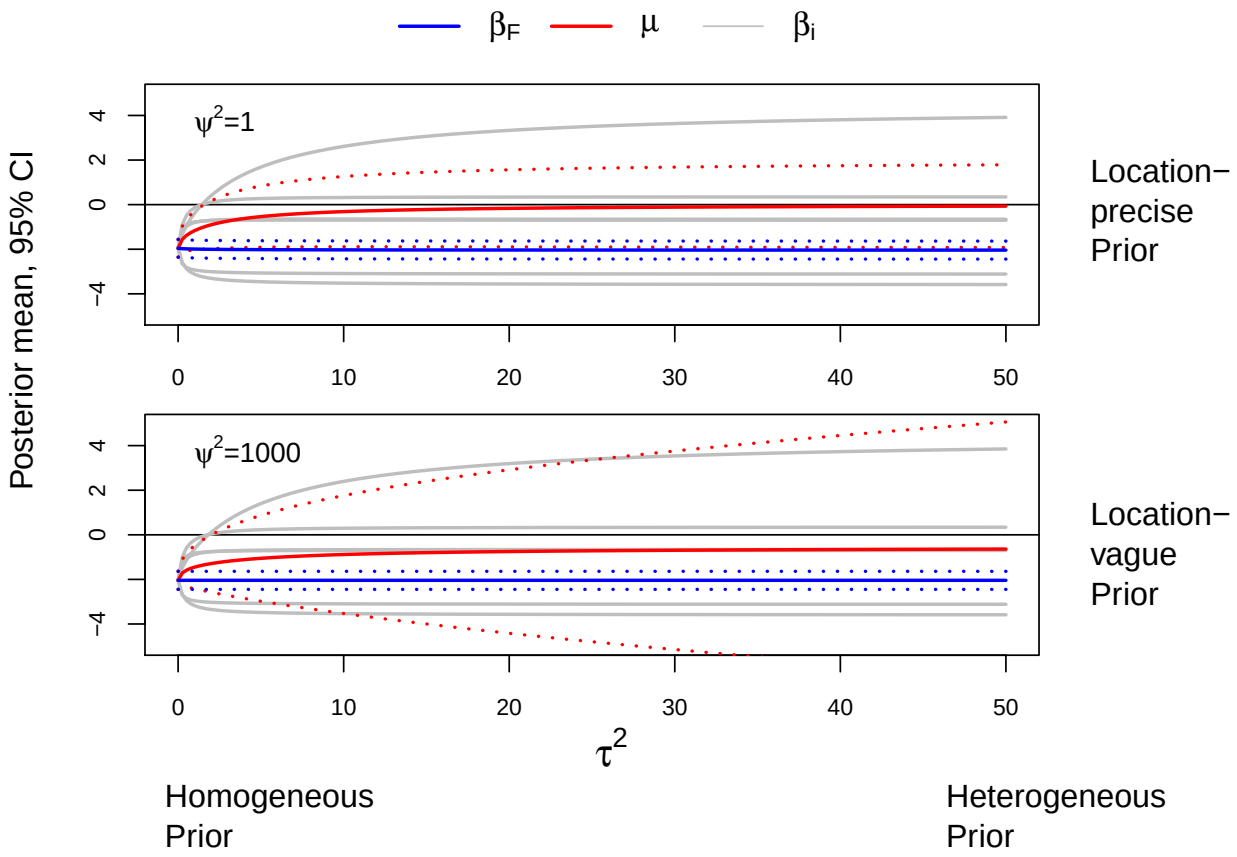


- Try ψ^2 at 0.1, 1, 1000
- Try τ^2 fixed at 1/4, 1, 25, 100 and a selection from Lambert (2005);

  $L_1 : \tau^{-2} \sim \Gamma(0.001, 0.001)$; $L_3 : \log(\tau^2) \sim \mathrm{U}(-10, 10)$; $L_5 : \tau^{-2} \sim \mathrm{U}(1/1000, 1000)$; $L_7 : \tau^{-2} \sim \mathrm{Par}(1, 0.001)$; $L_9 : \tau \sim \mathrm{U}(0, 100)$; $L_{11} : \tau \sim N(0, 100), \tau > 0$
- Priors matter for μ, not β_F

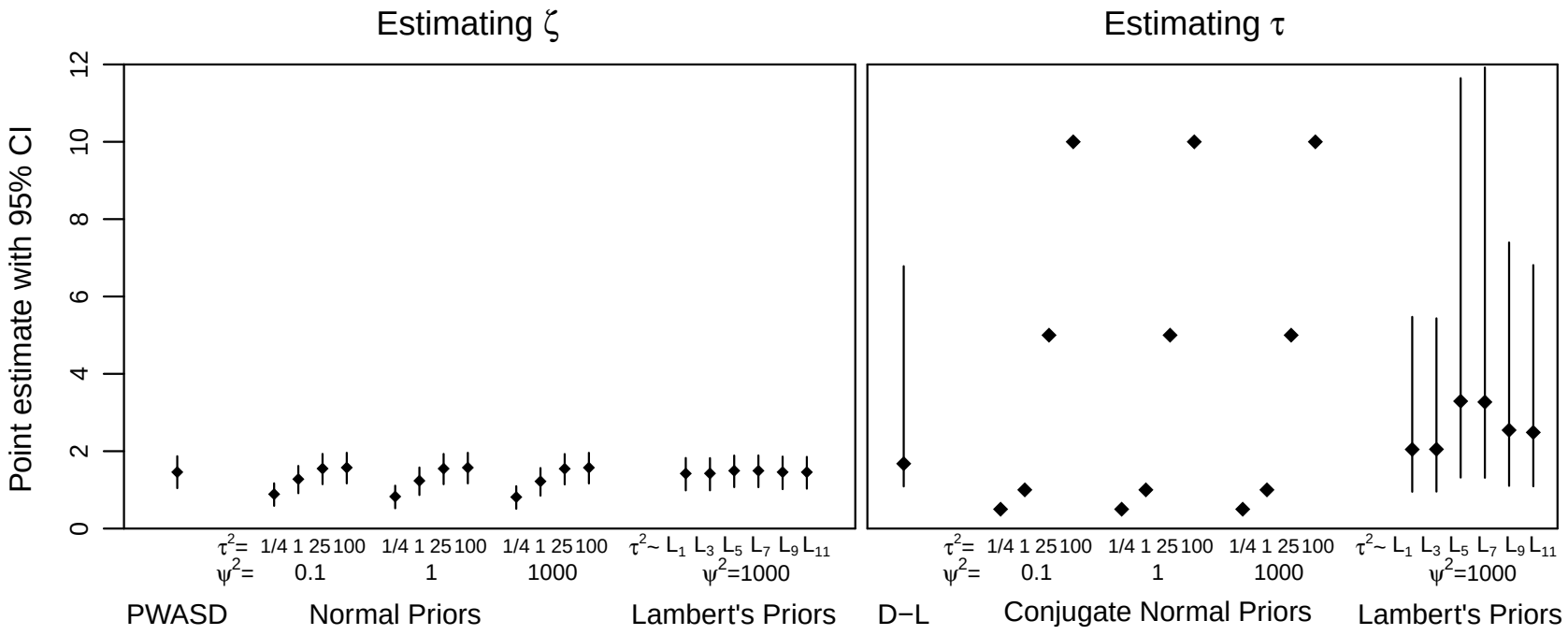
Meta-analysis: example

For the priors with fixed ψ, τ^2;



Meta-analysis: example

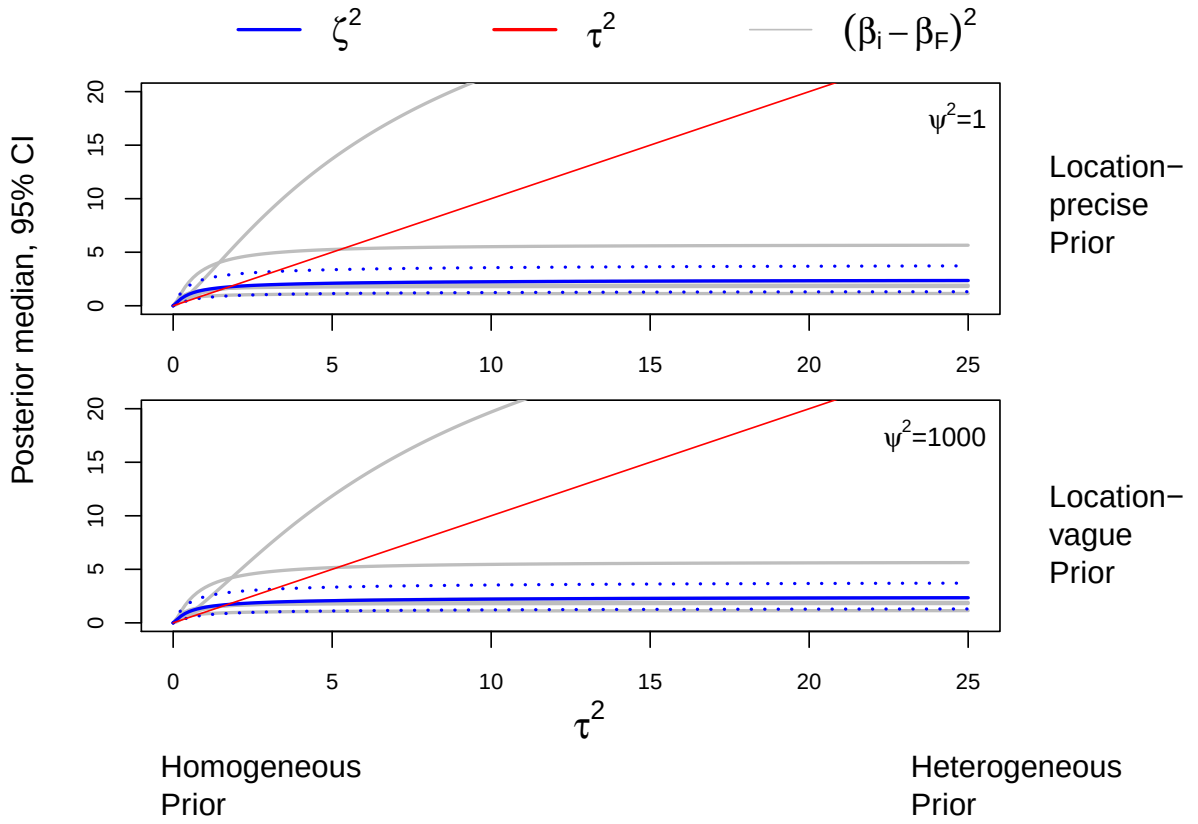
Similarly, precision-weighted 'spread' ζ^2 is more stable than τ^2



- Not as stable as for β_F – as data tell us less about ζ^2 than overall location
- Just reporting the original-data forest plot is a sane summary

Meta-analysis: example

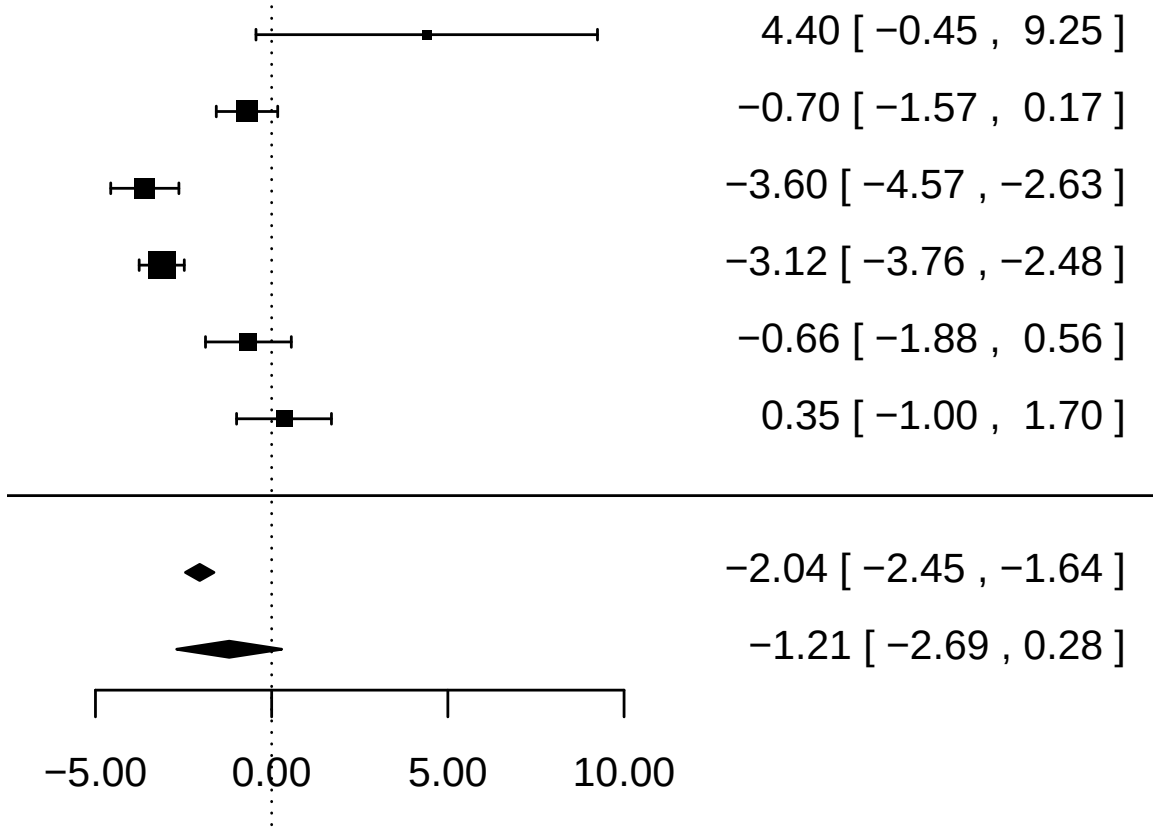
And for priors with fixed ψ^2, τ^2 – the same story;



More on sensitivity

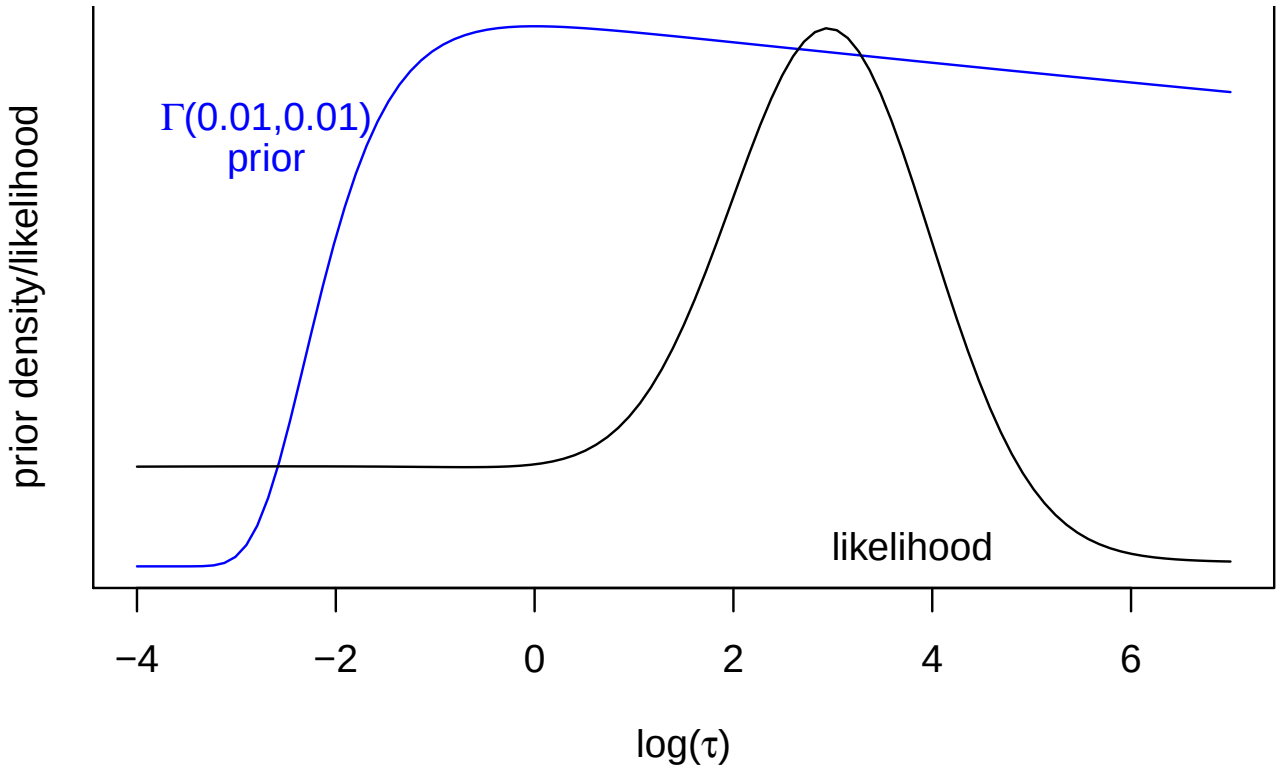
Why, in these models, does the prior on τ^2 matter so much?

Recall our example; what values of τ are plausible?



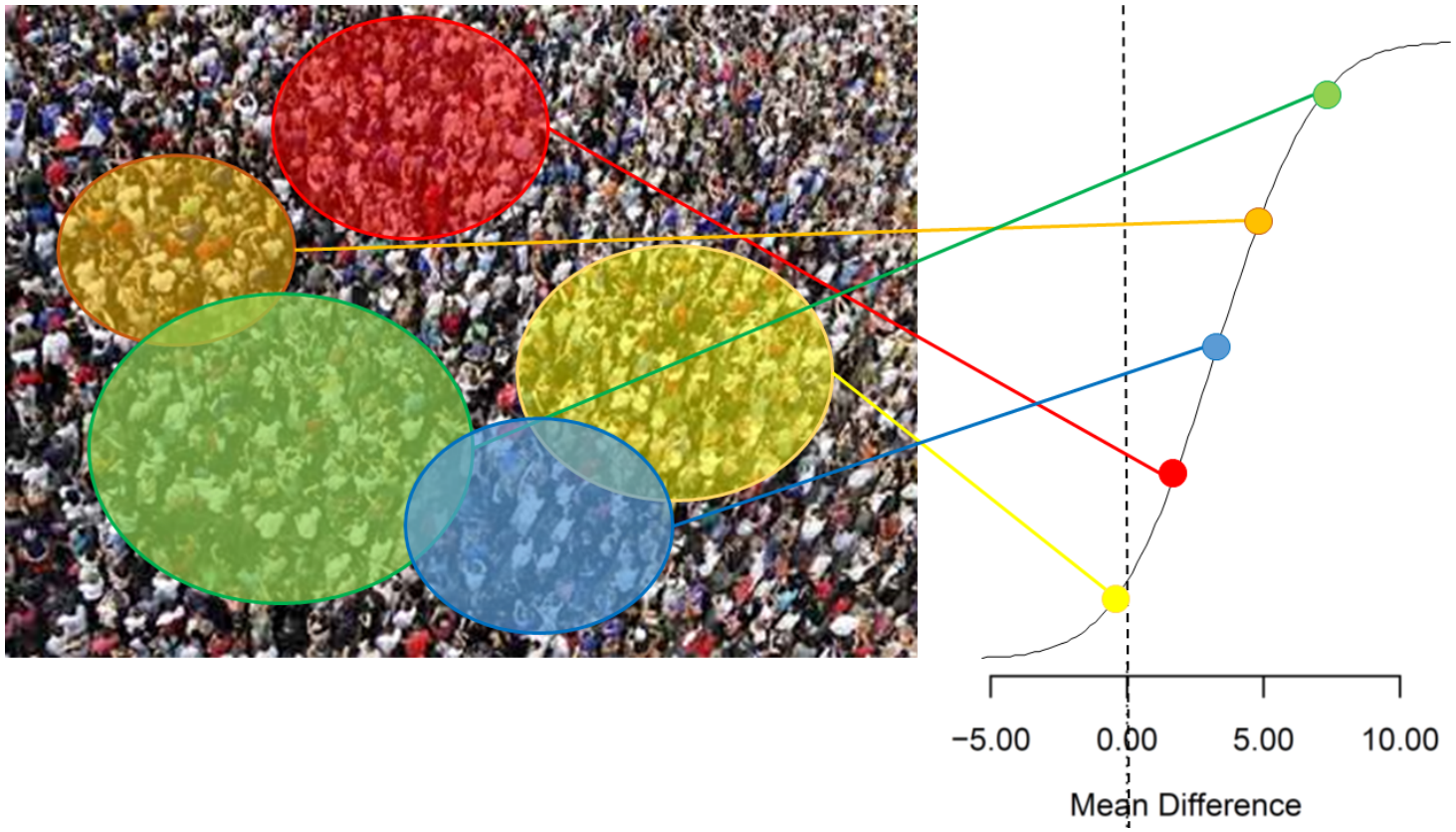
More on sensitivity

Homogeneity (i.e. $\tau = 0$) isn't ruled out by data – but low τ values **are**, under Γ priors. This can't be entirely avoided, expect to think carefully about the prior.



Hierarchical models: another motivation

Exchangeability is a strong justification for using hierarchical models (see e.g. Higgins & Spiegelhalter 2009). But the 'classical' motivation looks like this;



Randomly-sampled effect-sizes have mean μ, variance τ^2 – parameters of the *random effects* distribution.

Hierarchical models: another motivation

The same calculations can have >1 interpretation;

Model term			Random effects	Fixed effect**S** + exchangeability
$\widehat{\beta}_i$	$\sim$	$N(\beta_i, \sigma_i^2)$	Random outcomes	Random outcomes
β_i	$\sim$	$N(\mu, \tau^2)$	Random studies	Prior on fixed $\beta_1, ..., \beta_k$
μ	$\sim$	$N(0, \psi^2)$	Prior on fixed mean & var of possible β_i	
τ^2	$\sim$	$p(\tau^2)$		

- In RE model, ψ is the standard deviation of the prior on average study effect μ; τ is the standard deviation of the study effects
- An assumption of i.i.d. effects is often hard to justify; typically, later studies' designs depend on earlier studies' results – e.g. replication studies
- **But** random effects models are needed for prediction – what β_i might we see in the next study?

Hierarchical models: another motivation

Q. So will *this* upset people? A. Again (!) it depends who you ask

Random-effects models are unpopular with some...

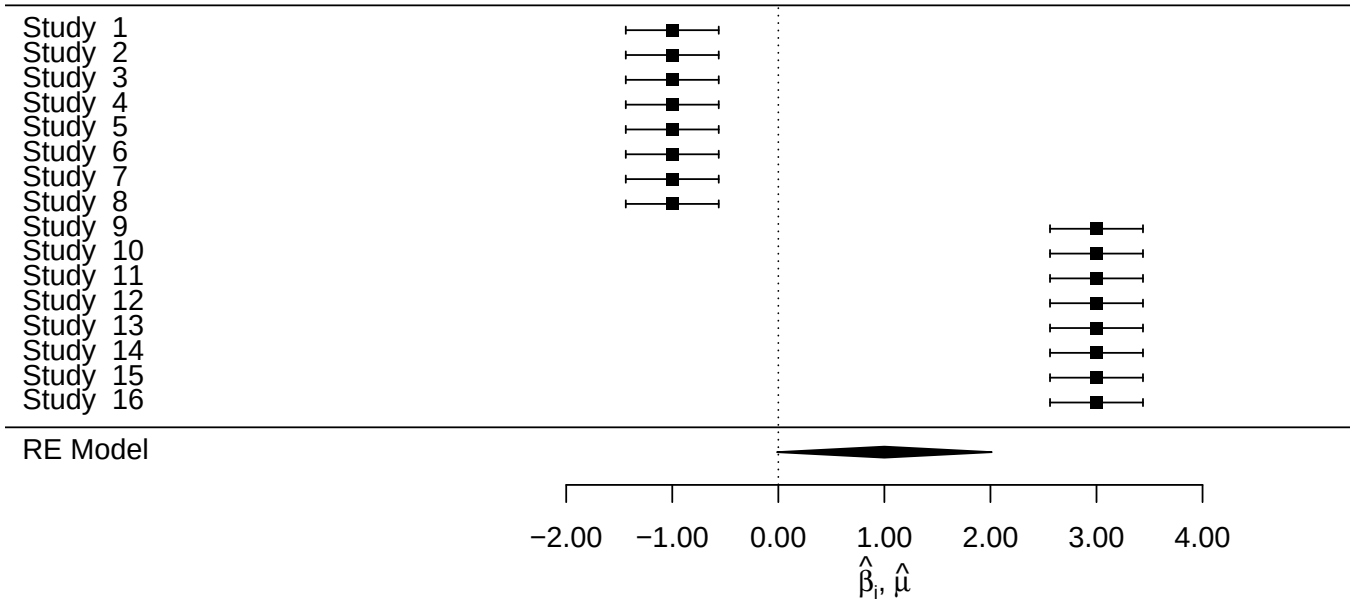
I'll not let the random differences between different trials contribute to my final p-value or contribute to my final estimate of the magnitude of the effect or to the confidence intervals that I'll put about it.

The random effects analysis says, look, we've got a lot of different trial results, here. What's the mean and what's the scatter of the different trials results? I don't think that this is actually wholly wrong [...] I think it does answer a question. But it's a very abstruse and uninteresting question

Richard Peto, Statistics in Medicine, 1987

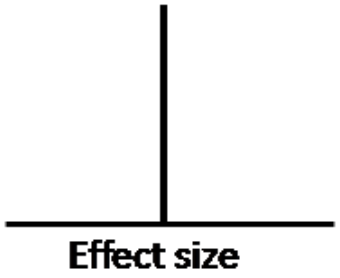
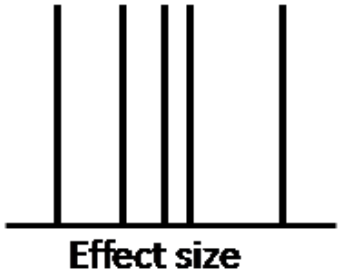
Hierarchical models: another motivation

It's worth noting that random-effects models do **not** provide intervals that 'reflect heterogeneity';



Recall that μ and its posterior describe the mean of the population of study effects *you might ever see*, not necessarily the set of effects in the observed studies.

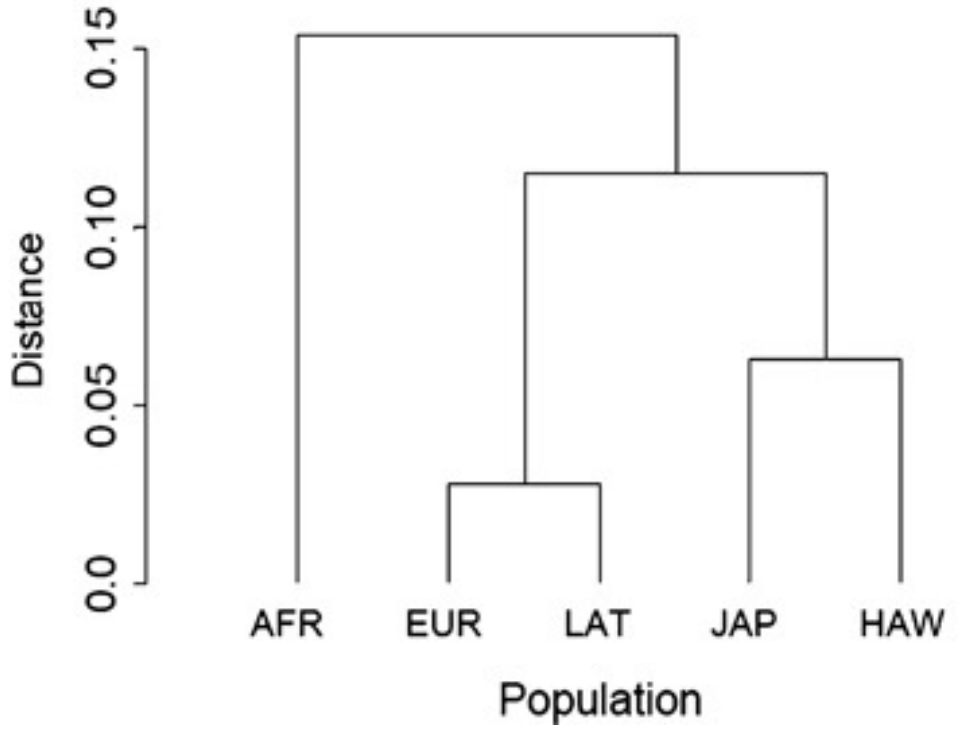
Summary – for inference

Name	Common effect	Fixed effect**S**	Random effects
			
Estimate:	β_0	β_F	μ
Spread:	nope!	ζ^2	τ^2
Problems?	Unrealistic	**Just right!**	Sensitive

MANTRA

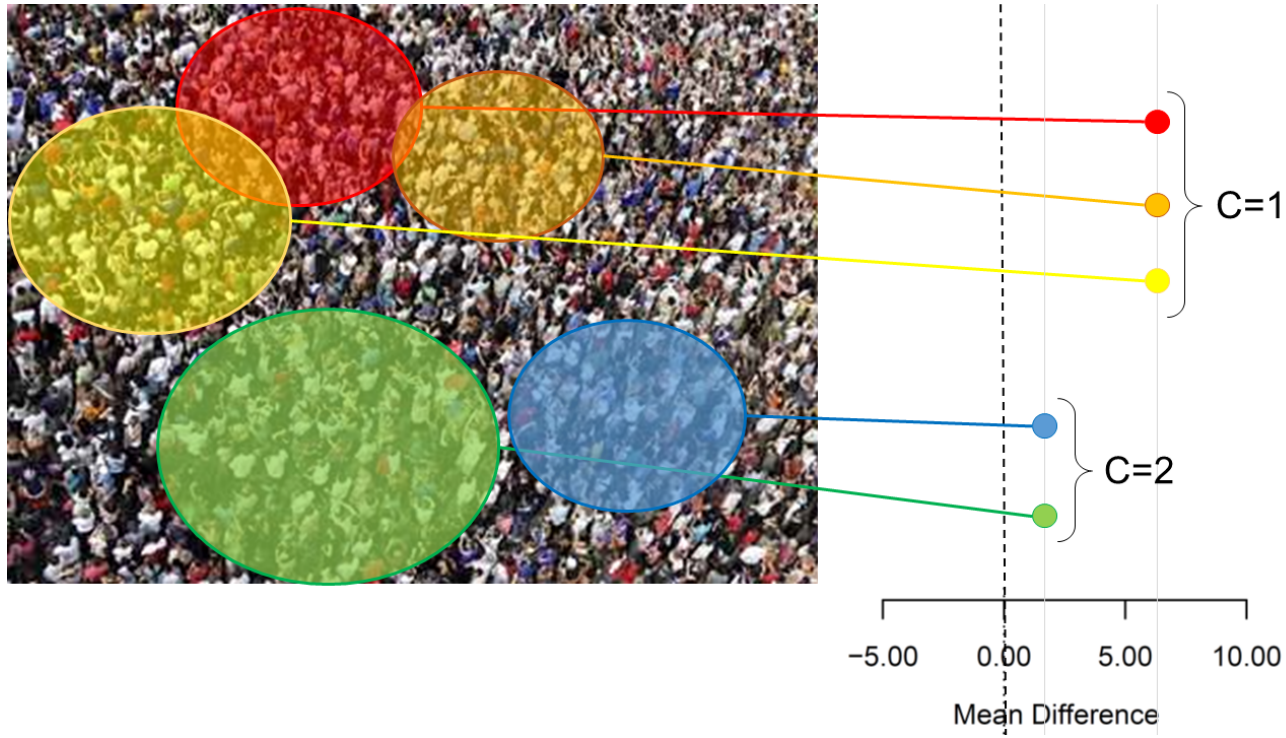
Assumptions of exchangeability provide attractive shrinkage and 'borrowing strength', when there's no reason to distinguish β_i. But, at least in genetic association work, ancestry may suggest which β_i may be similar;

Right: dendrogram illustrating mean MAF similarities/differences between African American, European American, Latinos, Japanese Americans & Native Hawaiians; data from the Type 2 Diabetes (T2D) consortium.



MANTRA

The Meta-ANalysis of Transethnic Association studies (MANTRA) method (Morris, 2011) exploits the MAF information (or F_{ST}) to cluster effects in sub-populations. Conceptually;



MANTRA

Within each cluster, there is a single 'center' effect selected from the study effects $\beta_1, \beta_2, \ldots \beta_k$ – each study is equally likely to be such a center, *a priori*, and each non-center study gets assigned to its 'nearest' center.

Within-cluster, the center effect size has prior

$$\begin{aligned} \beta_c &\sim N(\mu, \tau^2) \\ \tau^2 &\sim Exp(1) \\ \mu &\sim \text{flat.} \end{aligned}$$

The number of clusters C has prior

$$\mathbb{P}[C = c] = \begin{cases} \frac{1}{2}, & c = 1 \\ \frac{1}{2^c} \frac{2^{k-1}}{2^{k-1}-1}, & c = 2, ..., k, \end{cases}$$

i.e. homogeneity has 50% prior support, then it 'tails off'.

MANTRA

MANTRA is implemented with reversible jump MCMC – somewhat like Gibbs Sampling, but allowing center effects β_c to enter/leave the model. It is run twice, with all $\beta_i = 0$ (i.e. the null) and the model above (alternative).

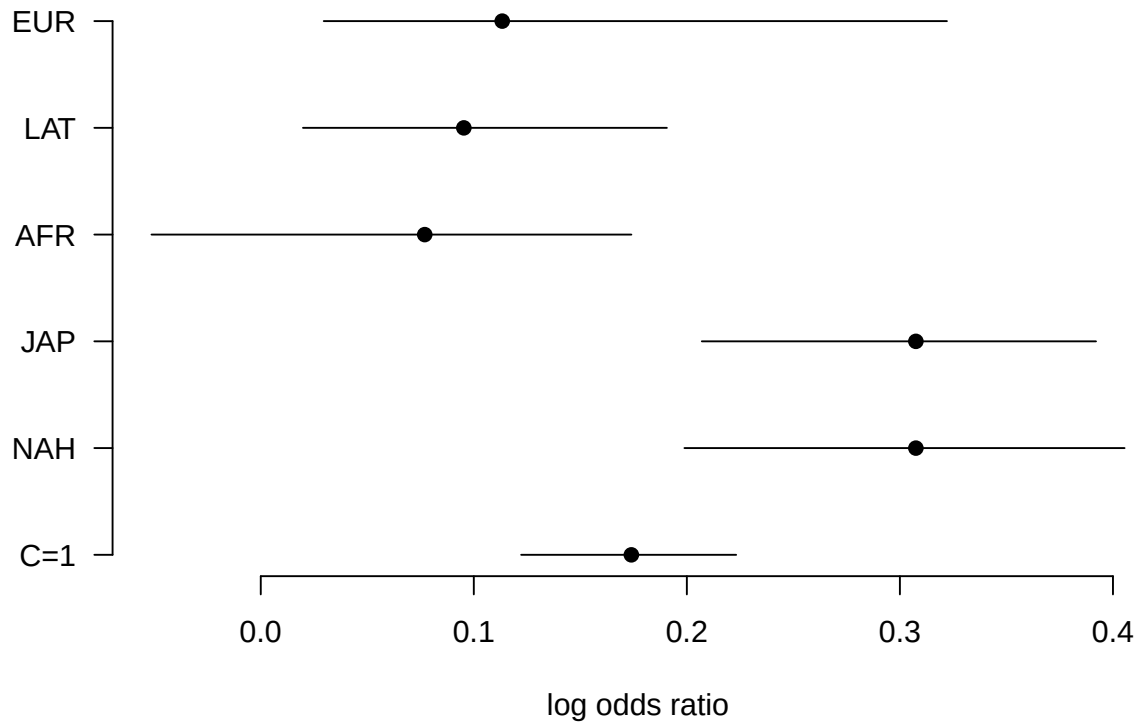
Its output;

- Bayes Factor comparing the null with the clustered, non-zero β_i
- Posterior probability of $C > 1$ under the alternative
- Posterior probabilities of cluster-membership, for each study, under the alternative

The computational effort required is non-trivial (e.g. 10 mins per SNP) but can be parallelized; 32 processors for 1 week enables GWAS with 2.5M SNPs.

MANTRA

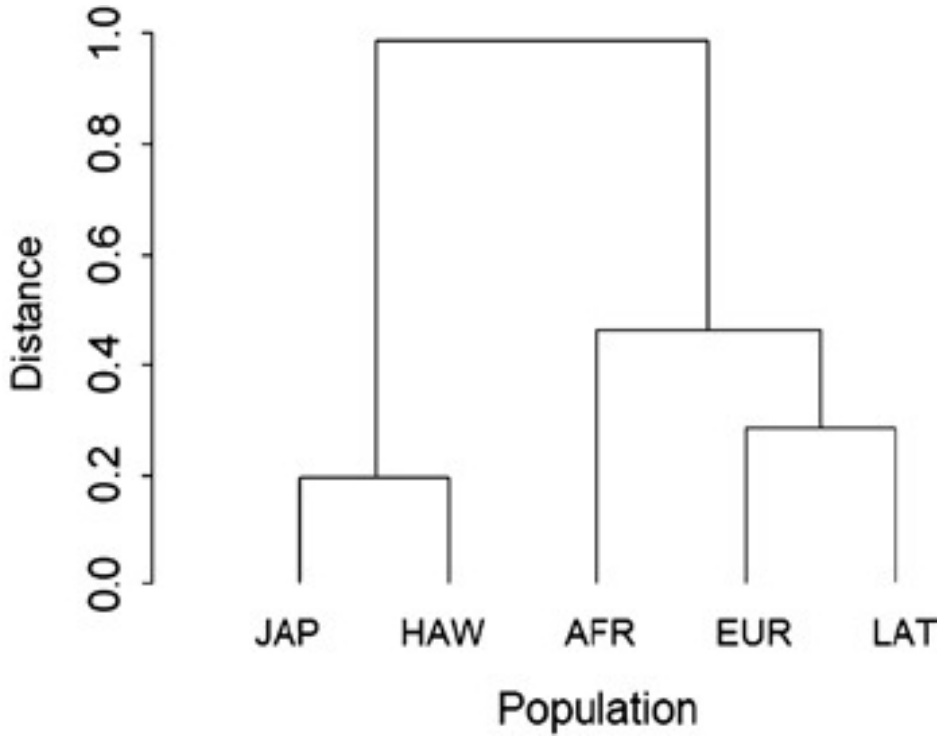
Output for T2D association, at rs7754840 in the (known) CDKAL1 locus;



Compared to the null, get $BF = 8.9$ for $C = 1$, but $BF = 11.0$ for unconstrained model – and 99.2% posterior probability that $C > 1$.

MANTRA

Showing the posterior probability of cluster memberships;



The big Bayes Factor occurs as the data suggest differences between group as well as a non-zero average effect. **Both** violate the null – that **all** $\beta_i = 0$.

MANTRA

Heterogeneity and average effect in the fixed-effects analysis; writing

$$\begin{aligned} Z_i^2 &= \hat{\beta}_i^2/\sigma_i^2 \\ Z_F^2 &= \hat{\beta}_F/\mathrm{Var}[\,\hat{\beta}_F\,], \\ \text{then } Z^2 &= \sum_{i=1}^{k} Z_i^2 \\ &= Z_F^2 + \sum_{i=1} \sigma_i^{-2}(\hat{\beta}_i - \hat{\beta}_F)^2 \\ &= Z_F^2 + Q, \end{aligned}$$

i.e. the signal-to-noise over all studies is the signal-to-noise for the average effect β_F plus the heterogeneity – Cochran's Q.

GWAS usually only examines β_F – but there's no need to restrict like this. See also the ASSET method, looking at differences by disease subtype.

Summary

- Meta-analysis is natural in a Bayesian framework
- Summarizing what You know is still a challenge
- Questions of heterogeneity are of interest, but often more sensitive to modeling assumptions; prior information matters

Obligatory XKCD cartoon

And finally;

The meta-*meta* analyses are real!