THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

# MODULE 1 | GENETIC MAPPING

## Session 7. Meta-analysis

July 2026

Slides, practicals & data can be downloaded from the cluster:
`/data/module1/downloadsTuesPM.zip`

Slides and prac guide can be downloaded from the website:
https://cnsgenomics.com/data/teaching/GNGWS26/module1/

*Why meta-analyse GWAS results?*

Sample size is a key consideration in GWAS for complex traits

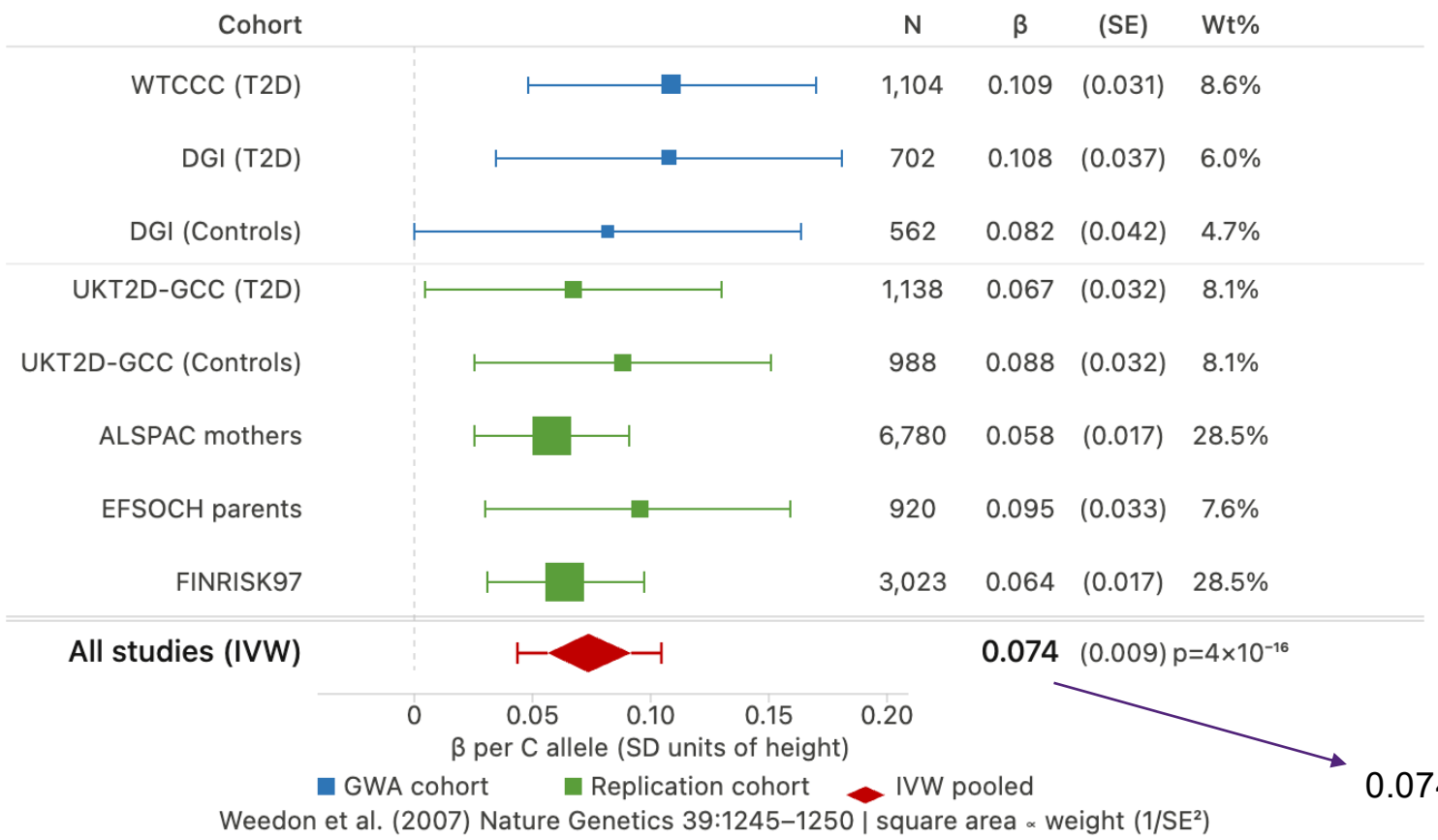
Is there heterogeneity between studies?

The process of combining results is called meta-analysis

Key technique underlying consortia

- Easier to share summary statistics than raw genotypes
- Little loss in statistical power – if a consistent model is applied

The basic idea of meta-analysis in GWAS is to calculate a ‘weighted mean’ per SNP




0.074 SD × 6.5 cm/SD ≈ **0.48 cm**

Most common approach is a fixed effect model using an **inverse-variance weighted method:**

- Estimates from each study are weighted by their precision (inverse of the variance for $\hat{\beta}$, or 1/SE$^2$)
- i.e. larger studies (with smaller SEs) have more weight

## Fixed-effect model

$$\beta = \frac{sum\ of\ (estimate\ \times\ weight)}{sum\ of\ weights} = \frac{\sum_k \beta_k w_k}{\sum_k w_k} \qquad se = \sqrt{\frac{1}{\sum_k w_k}}$$

$\beta_k$: effect estimate for study $k$

$w_k$: weight for study $i$, given as $\frac{1}{se_k^2}$

$se_k$: standard error for study $k$

If all the weights are the same, the weighted average is equal to the mean effect

The standard error can be used to derive:

- confidence interval: measure of precision (or uncertainty) of the summary estimate
- *P*-value: measure of strength of the evidence against the null hypothesis of no effect

*There is one true effect, and estimates from each study vary due to their precision*

Heterogeneity – does the true effect size varies between studies?

e.g. ancestry differences, GxE across cohorts, phenotype definition, adjustments, sex

**Cochran's Q statistic**

$$Q = \sum_k w_k(\widehat{\beta_k} - \widehat{\beta_{pool}})^2$$

$\widehat{\beta_k}$: effect estimate for study $k$

$\widehat{\beta_{pool}}$ = pooled fixed effect estimate

$w_k$: weight for study $i$, given as $\frac{1}{se_k^2}$

$se_k$: standard error for study $k$

- Under the null hypothesis of a common effect size, Q follows a chi-squared distribution with *K - 1* degrees of freedom. Low p-value ~ high heterogeneity.
- Limitations:
  - low power with few studies
  - for many studies, even small heterogeneity becomes significant

Heterogeneity – does the true effect size varies between studies?

e.g. ancestry differences, GxE across cohorts, phenotype definition, adjustments, sex

**$I^2$ statistic**

$$I^2 = \frac{Q-(K-1)}{K}$$

$Q$: Cochran's Q
K = total number of studies

- The $I^2$ statistic describes the proportion of variation across studies that is due to heterogeneity rather than chance
- Unlike Q, not inherently dependent on number of studies
- $I^2 > 0.50$ – consider a random effects model

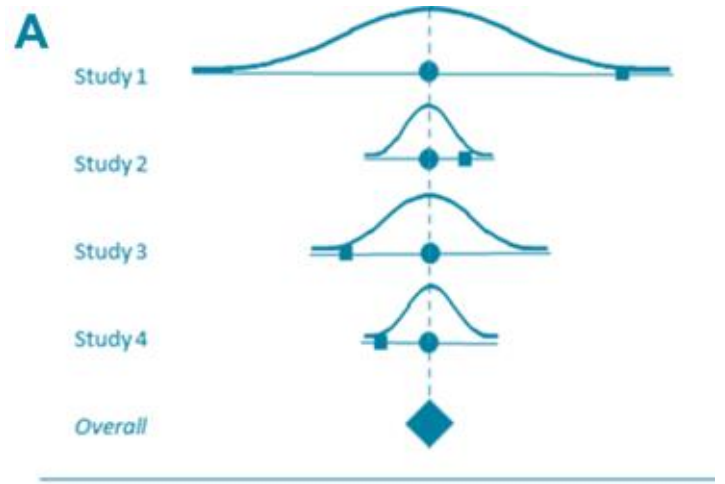
## Random-effect model

Random-effects model assumes that the true effect across studies are drawn from a normal distribution, i.e. they are different between studies.

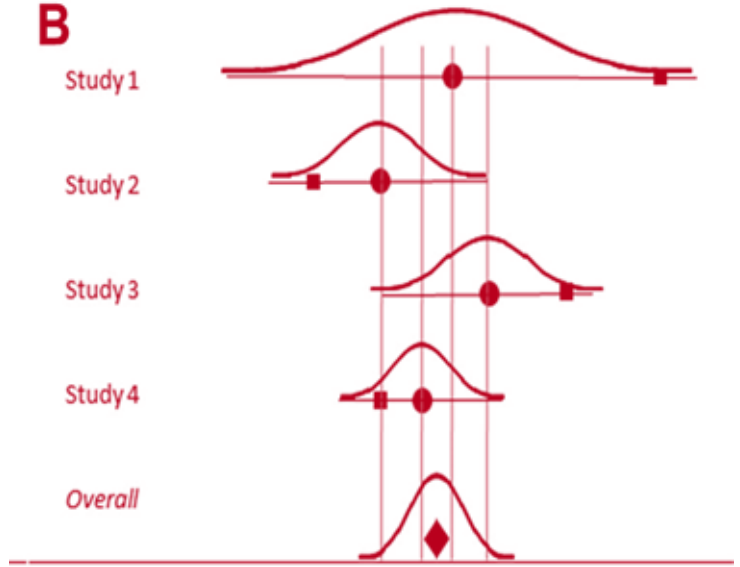
Two sources of variance – within studies plus between studies. Weights less dependent on sample size.

The overall (meta-analysed) effect is the mean of the normal distribution.

Fixed-effect:




Random-effect:




Spineli 2020 Am J Orthod Dentofacial Orthop.

Meta-analysis of large cohorts requires meticulous QC and planning...

1. **Trait definition**

   The trait should be the same across studies, same covariates adjustments and transformations.

2. **Analysis and QC pipeline (within studies)**

   Consistent QC pipeline (e.g. Hardy-Weinberg equilibrium, genotype missing rate, imputation scores), DNA strand flipping (!), analysis model

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Case A: unambiguous e.g. C/T SNP | Forward strand: | 5’ | T | A | C | T | G | 3’ |
| | Reverse strand: | 3’ | A | T | G | A | C | 5’ |
| | Forwards strand: | 5’ | T | A | T | T | G | 3’ |
| | Reverse strand: | 3’ | A | T | A | A | C | 5’ |
| Case B: Ambiguous, palindromic SNPs e.g. C/G SNP | Forward strand: | 5’ | T | A | C | T | G | 3’ |
| | Reverse strand: | 3’ | A | T | G | A | C | 5’ |
| | Forwards strand: | 5’ | T | A | G | T | G | 3’ |
| | Reverse strand: | 3’ | A | T | C | A | C | 5’ |

Meta-analysis of large cohorts requires meticulous QC and planning...

1. **Trait definition**

   The trait should be the same across studies, same covariates adjustments and transformations.

2. **Analysis and QC pipeline (within studies)**

   Consistent QC pipeline (e.g. Hardy-Weinberg equilibrium, genotype missing rate, imputation scores), DNA strand flipping (!), analysis model

3. **Independence of the samples**

   is any relatedness between participants across studies?

4. **Meta-analysis approach**

   Fixed, random or something else

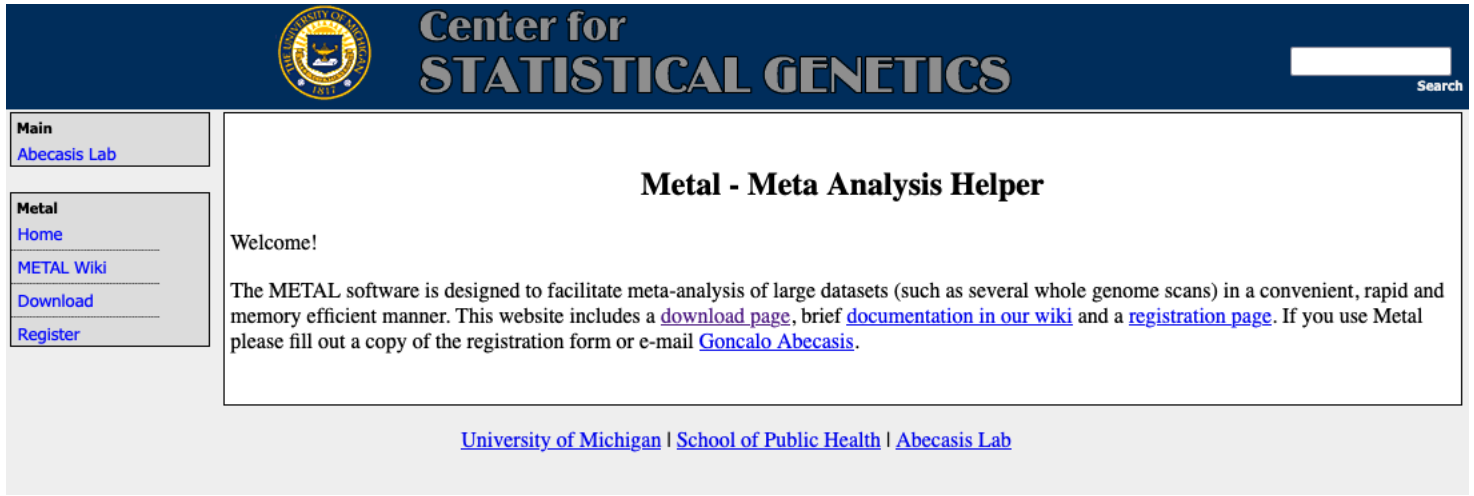
   What do you plan to do with the results? what output do you require?

1. **PLINK – used in the practical**
2. **Metal – most common**
3. **Many other options & variations**

   Fisher's approach, p-values only

   Across traits, correlated phenotypes

   Bayesian methods



## Links for further reading

- Cochrane Training Chapter 10:  Analysing data and undertaking meta-analyses
- Doing Meta-Analysis with R: A Hands-On Guide
- Evangelou et al. 2013 Nat Rev Genet
- Zeggini 2009 Pharmacogenomics