MODULE 1 | GENETIC MAPPING

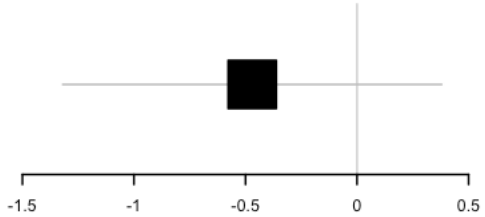
Session 8. Meta-analysis

20 June 2023

- Sample size is a key consideration in GWAS
- Small samples have low power to detect associations, particularly at the genome-wide significance threshold of $P < 5 \times 10^{-8}$
- So, how can we derive information from small GWAS? One option is to meta-analyse with other studies, i.e. combine association results across studies
- As we will see in this lecture, there are several advantages of using meta-analysis in genomics research, but there are also important considerations to be taken

- Consider this toy example. What would you conclude from these results?
- What if we add information from an independent study, with the same number of participants and same effect estimates?

Study	N	Estimate	SE	P
Study 1	1000	-0.47	0.4333	0.278

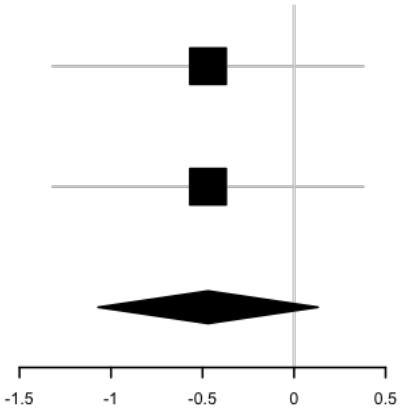


Based on this study, there is no evidence that the variant is associated with the trait studied

Example adapted from Cochrane Training

- Consider this toy example. What would you conclude from these results?
- What if we add information from an independent study, with the same number of participants and same effect estimates?

Study	N	Estimate	SE	P
Study 1	1000	-0.47	0.4333	0.278
Study 2	1000	-0.47	0.4333	0.278
Summary	**2000**	**-0.47**	**0.3064**	**0.125**

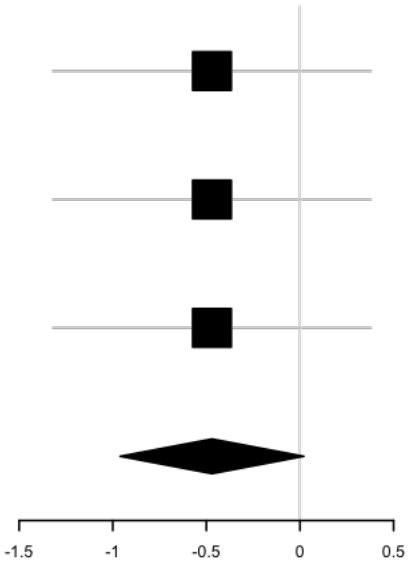


The added evidence increased confidence in the estimate, but error is still large

Example adapted from Cochrane Training

- Consider this toy example. What would you conclude from these results?
- What if we add information from an independent study, with the same number of participants and same effect estimates?

Study	N	Estimate	SE	P
Study 1	1000	-0.47	0.4333	0.278
Study 2	1000	-0.47	0.4333	0.278
Study 3	1000	-0.47	0.4333	0.278
Summary	**3000**	**-0.47**	**0.2502**	**0.06**

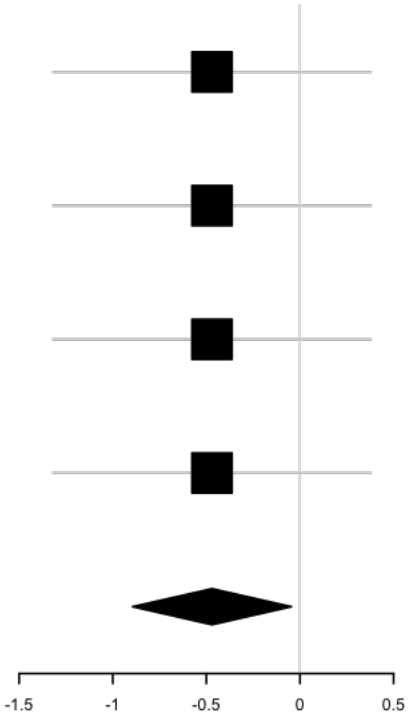


With a 3rd study added the error of the estimate decreases further, and the probability of seeing this effect by chance (i.e. if there is no true effect in the population) is smaller

Example adapted from Cochrane Training

- Consider this toy example. What would you conclude from these results?
- What if we add information from an independent study, with the same number of participants and same effect estimates?
- Note that the point estimate did not change, only its precision
- This is an extreme example. In real life we see ≠ estimates across studies with ≠ SEs, depending on the real effect in the population.

Study	N	Estimate	SE	P
Study 1	1000	-0.47	0.4333	0.278
Study 2	1000	-0.47	0.4333	0.278
Study 3	1000	-0.47	0.4333	0.278
Study 4	1000	-0.47	0.4333	0.278
Summary	**4000**	**-0.47**	**0.2167**	**0.03**



With a 4th study added the P-value is now nominally significant

Example adapted from Cochrane Training

What are key advantages of using meta-analysis in genomics research?

1. **Increase statistical power**

   As we saw in the toy example, small studies may have insufficient power to identify true genetic effects. However, combining information from independent studies can improve precision of the effect estimates. This is particularly relevant when it comes to detecting subtle genetic effects.

2. **Increase sample size without sharing individual-level data**

   Sharing individual-level genetic data across research groups raises privacy concerns. Meta-analyses overcome those issues and other ethical considerations by relying solely on summary statistics.

3. **Identify heterogeneity across studies**

   Meta-analysis provide the opportunity to investigate potential sources of heterogeneity (e.g., study design, population characteristics, or genotyping methods).

4. **Resolving inconsistent findings**

   Inconsistent or contradictory results across individual studies can be explored through meta-analysis. Meta-analysis can help identify the sources of discrepancies, evaluate the overall effect size, and provide a more accurate assessment of the true association.

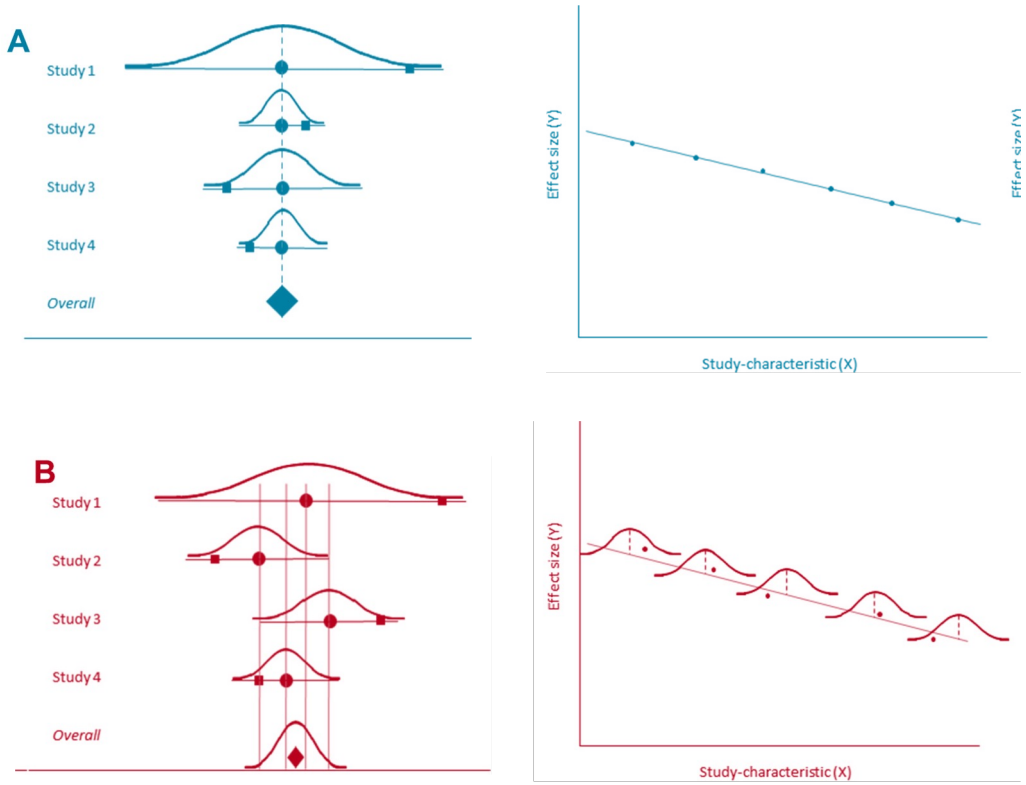
A common and simple approach to conduct meta-analysis is to use an **inverse-variance weighted method:**

- Estimates from each study are weighted by the inverse of the variance of the effect estimate ($1/SE^2$)
- Larger studies (with smaller SEs) are given more weight

Fixed vs. random effects models

- **Fixed-effect model:** assumes that the true genetic effect being estimated is the same across studies, and that observed variations between studies are caused by chance.

- **Random-effects model:** assumes that the true genetic effects being estimated across the different studies are different, yet related, i.e. this model assumes substantial diversity in effect estimates and assesses both intra-study sampling errors and inter-study variances.



Spineli 2020 Am J Orthod Dentofacial Orthop.

A common and simple approach to conduct meta-analysis is to use an **inverse-variance weighted method:**

- Estimates from each study are weighted by the inverse of the variance of the effect estimate ($1/SE^2$)
- Larger studies (with smaller SEs) are given more weight

Fixed-effect model

$$\beta = \frac{sum\ of\ (estimate \times weight)}{sum\ of\ weights} = \frac{\sum_i \beta_i w_i}{\sum_i w_i} \qquad se = \sqrt{\frac{1}{\sum_i w_i}}$$

β_i: effect estimate for study i

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

se_i: standard error for study i

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

Fixed or random effects model?

"*The choice of meta-analysis model depends on the presence or absence of heterogeneity. In the absence of heterogeneity, a fixed effects model is used for meta-analysis.*" (Lee 2015 Ann Lab Med)

Heterogeneity measures

- $\boldsymbol{I^2}$ ***statistic*****:** percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

Q: χ^2 statistic
df: χ^2 degrees of freedom

Rough interpretation guide:

0% to 40%: might not be important
30% to 60%: may represent moderate heterogeneity
50% to 90%: may represent substantial heterogeneity
75% to 100%: considerable heterogeneity

↓ BUT…

Importance of I^2 depends on:

- magnitude and direction of effects
- strength of evidence for heterogeneity (e.g. *P*-value from the χ^2 test, or CI for I^2)

https://training.cochrane.org/

Fixed or random effects model?

"*The choice of meta-analysis model depends on the presence or absence of heterogeneity. In the absence of heterogeneity, a fixed effects model is used for meta-analysis.*" (Lee 2015 Ann Lab Med)

Heterogeneity measures

- $\boldsymbol{I^2\ statistic}$: percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

Q: χ^2 statistic
df: χ^2 degrees of freedom

- When N is small, non-significant ≠ evidence of no heterogeneity
- Heterogeneity more likely when meta includes more studies

BUT…

- $\boldsymbol{Cochran's\ Q\ statistic}$: indicates the presence of statistical heterogeneity. This is a χ^2 test that assesses whether observed differences in results are compatible with chance alone. A low *P*-value (or a large χ^2 statistic relative to its df) provides evidence of heterogeneity of effects (variation in effect estimates beyond chance).

https://training.cochrane.org/

Some important considerations for GWAS meta-analyses

1. **Trait definition**

   Ideally, trait definitions should be the same across studies and same covariates adjustments used. Similarly, it is important to consider any data transformations (scale of the effects).

2. **Quality checks**

   Have individual studies used appropriate QC (e.g. Hardy-Weinberg equilibrium, genotype missing rate, imputation scores)?

3. **Heterogeneity**

   Heterogeneity in effect size estimates may come from several sources. Phenotype variability may cause heterogeneity and may result in spurious associations. Heterogeneity due to ≠s in ancestry can occur given differences in LD with true causal variants. Other differences between studies (e.g. genotyping platforms, imputation software, QC, etc.) can also introduce heterogeneity.

4. **Independence of the samples**

   It is very important to consider if there is any relatedness between participants across studies as this can bias results.

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