Measuring Heterogeneity

Cochran's Q

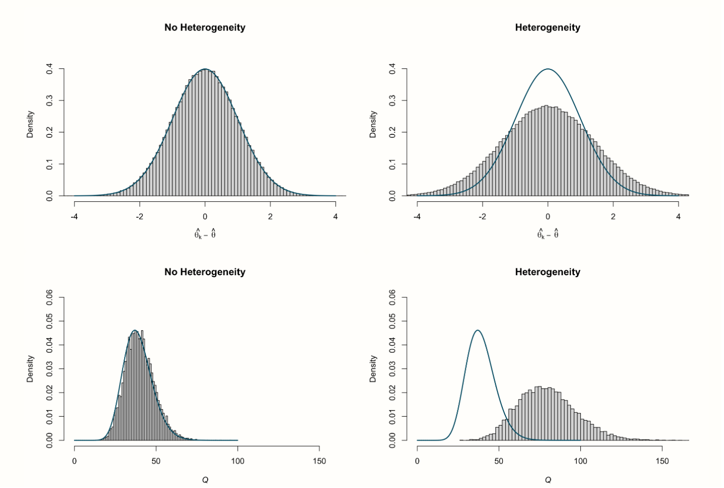
- Traditionally, Cochran's Q (Cochran, 1954) has been used to distinguish studies' sampling error from actual between-study heterogeneity.
- Q uses the deviation of each study's observed effect $\hat{\theta}_k$ from the summary effect $\hat{\theta}$ weighted by the inverse of the study's variance, w_k:

$$Q = \sum_{k=1}^{K} w_k \left(\hat{\theta}_k - \hat{\theta}\right)^2$$

- We assume that, if variability in observed effect sizes is only caused by sampling error (i.e., the EE-model holds), then Q will follow a χ^2 distribution with $K-1$ degrees of freedom, where K is the number of studies.

$$Q \sim \chi^2_{K-1}$$

→ If Q exceeds this expected value, we are more and more likely to conclude "significant" excess variability (=heterogeneity).



Measuring Heterogeneity

Cochran's Q

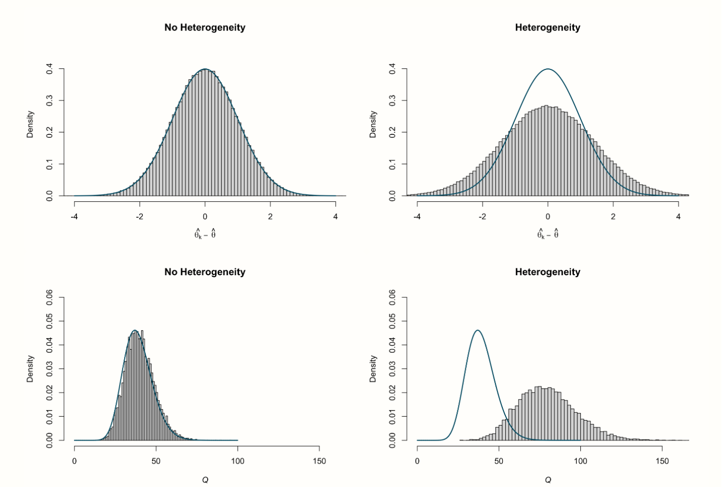
Problems With Q & the Q-Test

- Q increases both when the number of studies K, and when the precision (i.e. the sample size of a study) increases.

- Therefore, Q and whether it is significant highly depends on the **size of the meta-analysis**, and thus its **statistical power**.

→ **We should not only rely on the significance of a Q-test when assessing heterogeneity.**

→ Sometimes, meta-analysts decide whether to apply a fixed-effect or random-effects model based on the significance of the Q-test. This is also highly discouraged



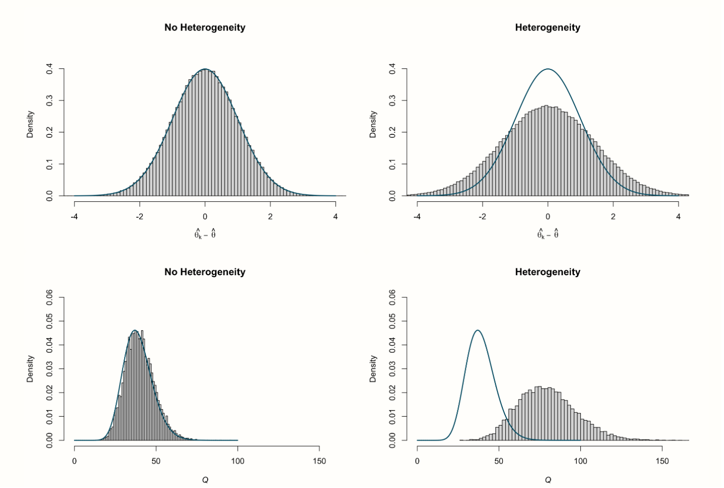
Measuring Heterogeneity

Higgins and Thompson's I^2

- I^2 is another metric to quantify the between-study heterogeneity in meta-analyses
- In its "classic" definition, I^2 quantifies, in percent, **how much the observed value of Q exceeds the expected Q value when there is no heterogeneity** (i.e., $K-1$):

$$I^2 = \max\left\{0, \frac{Q-(K-1)}{Q}\right\}$$

- The popularity of this statistic may be associated with the fact that there is an (in)famous "rule of thumb" how to interpret it (Higgins & Thompson 2002):
  - I^2 = 25%: low heterogeneity
  - I^2 = 50%: moderate heterogeneity
  - I^2 = 75%: substantial heterogeneity.



Measuring Heterogeneity

Higgins and Thompson's I^2

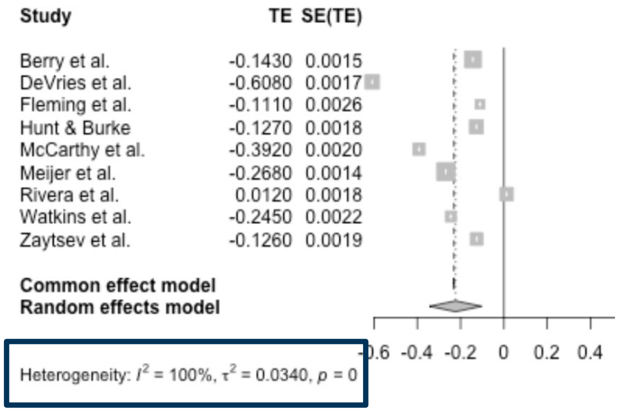
Problems of I^2

- I^2 is <u>not</u> sensitive to changes in the **number of studies** in the analysis.

- However, it is still a **relative measure of heterogeneity**, and its value heavily **depends on the precision** of the included studies

- I^2 is the **percentage of variability not caused by sampling error**.

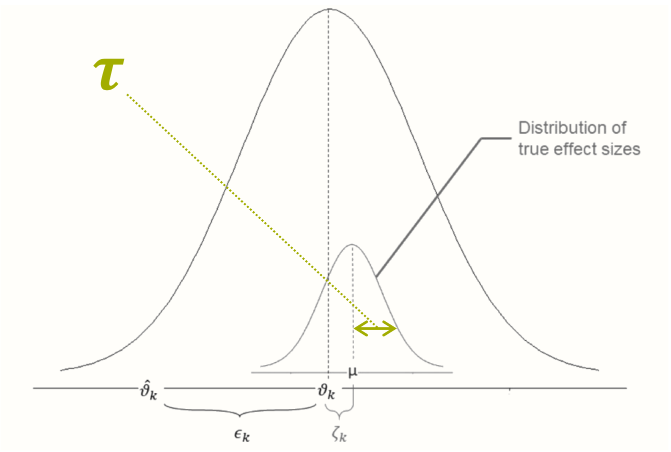
- If our studies become increasingly large, the sampling error tends to zero, while at the same time, $\boldsymbol{I^2}$ **tends to 100%** – simply because the **studies have a greater sample size**.



$$I^2 = \max\left\{0, \frac{Q-(K-1)}{Q}\right\}$$

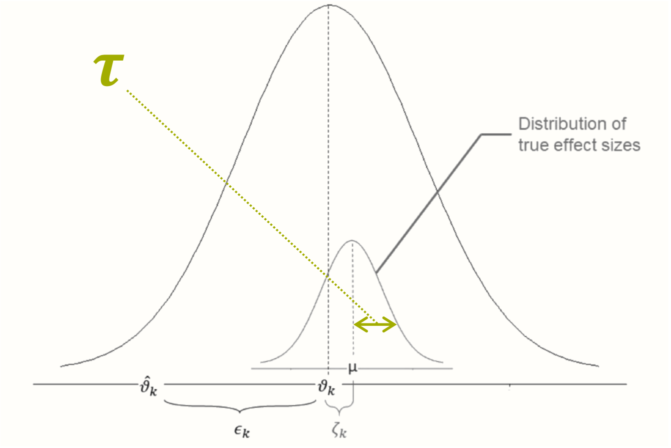
Measuring Heterogeneity

- Only τ^2 is **insensitive** to both the **number** and **precision** of the included studies
- τ (square root of τ^2) can be interpreted as the **standard deviation of the true effect size distribution**
- But this is often **difficult to communicate** to stakeholders without training in meta-analysis



Measuring Heterogeneity

- **Solution:** calculate **prediction intervals (PI)**, which directly express the impact that τ^2 has on the scale of the effect size
- Indicates the interval into which effect sizes of new studies are expected to fall, based on present evidence:



Upper PI: $\hat{\mu} + t_{K-1,0.975}\sqrt{\widehat{\mathrm{SE}}_{\hat{\mu}}^2 + \hat{\tau}^2}$

Lower PI: $\hat{\mu} - t_{K-1,0.975}\sqrt{\widehat{\mathrm{SE}}_{\hat{\mu}}^2 + \hat{\tau}^2}$

- PIs can be much wider than the CI.
- As $\tau^2 \rightarrow 0$, the PI and CI become increasingly identical.

Outliers & Model Diagnostics

- It is possible that **very large studies**, or studies with **very extreme effect sizes** may dominate the overall effect
- It is advisable to run sensitivity analyses excluding such studies
- There are **no iron-clad rules** on how to define outliers in a meta-analysis (but Viechtbauer & Cheung, 2010 provide helpful diagnostics)
- An intuitive way to identify outliers is through a **Baujat plot**