Methods to describe treatment effect heterogeneity in individual patient data meta-analysis

Karla Hemming, Jim Hughes, Andrew Forbes, Jo McKenzie

University of Birmingham; Washington University; Monash University

Cochrane Methods Group

May 7, 2021
Meta-analysis

- Conventional meta-analysis based on summary level data\textsuperscript{1}:
  - For every study an estimate of the treatment effect (SD) is available;
  - These treatment effects are pooled to obtain a single summary estimate of the treatment effect (along with CI).

- IPD meta-analysis using individual level data\textsuperscript{2}:
  - These data are pooled either using a two-stage approach (above);
  - Or, the data are analyzed using a one-stage approach using a generalised linear mixed model (details later).

- Two common approaches, fixed or random:
  - Fixed effect meta-analysis (common treatment effect across studies);
  - Random effect meta-analysis (heterogeneity of treatment effects).

\textsuperscript{1}Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ. 1997

Treatment effect heterogeneity in meta-analysis

- Test-statistic for treatment effect heterogeneity:
  \[ Q = \sum w_i(\hat{\mu}_i - \hat{\mu}_F)^2 \]  
  where \( i \): study; \( \hat{\mu}_i \): treatment effect for study \( i \); \( w_i \): precision for study \( i \); \( \hat{\mu}_F \): weighted pooled estimate.

- If \( w_i^{-1} = \hat{\sigma}^2 \) (variance of the treatment effect for study \( i \)) does not vary across studies the intuitive measure of between study heterogeneity is:
  \[ \hat{I}^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2} = \frac{\text{between}}{\text{between} + \text{within}} \]  
  where \( \hat{\tau}^2 \) is the (estimated) variance of the distribution of the \( \mu_i \)'s across the studies.

- If \( w_i \) is allowed to vary across studies it turns out that:
  \[ \hat{I}^2 = 100 \times \frac{Q - (K - 1)}{Q} \]  
  where \( K \) is the number of studies.
Finally, the predictive interval, which represents a region in which it is expected that 95% of future trial specific treatment estimates will fall, will be:

\[
[\hat{\mu}_F - t_{\alpha/2,K-1} \sqrt{((\hat{\tau}^2 + SE(\hat{\mu}_F)^2)} \text{ to } \hat{\mu}_F + t_{\alpha/2,K-1} \sqrt{((\hat{\tau}^2 + SE(\hat{\mu}_F)^2)}]
\]  

(4)
Simulated IPD data (true I-squared 91%)

10 trials; High treatment effect heterogeneity (I-squared = 87.2%; Q-statistic=70.13 (d.f=9) p=0.000; $\hat{\tau}^2 = 0.14$); two-stage approach.

We will go on to consider how this analysis could be conducted using a one-stage approach.
Linear mixed model with treatment effect heterogeneity

Model treatment effect heterogeneity using an “interaction” term and allowing for a covariance term:

\[ y_{ijs} = \beta_0 + x_{ijs}\theta + \alpha(S)_j + x_{js}\alpha(ST)_j + e_{ijs} \]  

\[ i : \text{individual}; j : \text{study}; s : \text{arm} \ (m \ per \ arm) \]

and that

\[
\begin{pmatrix}
\alpha(S)_j \\
\alpha(ST)_j
\end{pmatrix}
\sim
N\left(
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\tau^2_S & \sigma^2_{ST} \\
\sigma^2_{ST} & \tau^2_{ST}
\end{pmatrix}
\right)
\]

\[ S : \text{Study effect}; ST : \text{Study by Treatment effect} \]

Relationship to MA

\[ \tau^2_{ST} \] represents the variation between studies in their response to treatment (and so is akin to \( \tau^2 \) in a meta-analysis).

Proposed I-squared one-stage

Recap: I-squared

The between study variability of the treatment effect divided by the sum of the between-study variability and the within-study variability

When analysing using a one-stage approach an intuitive estimate of I-squared is thus:

\[ I^2 = \frac{\tau_{ST}^2}{\tau_{ST}^2 + \frac{2\sigma_e^2}{\bar{m}}} \]

\( \bar{m} : \text{average (harmonic mean) study size per arm} \)

Prediction interval:

\[ [\hat{\theta} - t_{\alpha/2, K-1} \sqrt{(\hat{\tau}_{ST}^2 + SE(\hat{\theta})^2)} \text{ to } \hat{\theta} + t_{\alpha/2, K-1} \sqrt{(\hat{\tau}_{ST}^2 + SE(\hat{\theta})^2)}] \] (6)
Recap: Simulated IPD meta-analysis

10 trials; High treatment effect heterogeneity (True I-squared = 91%; Q-statistic=70.13 (d.f = 9) \( p = 0.000; \hat{\tau}^2 = 0.14 \));

I-squared

Using one-stage approach \( \hat{I}^2 \) is 91.7% (87% using two-stage approach).
Simulation study

- Performance measures: correlation and bias;
- \( N = 10,000 \) data-sets simulated for each scenario;
- Data simulated from a linear mixed model with random study and random study by treatment interaction.
- Scenarios considered: 108:
  - REML and DL methods;
  - Number of studies: 10, 50, 100;
  - Study size per arm: 10, 50, 100;
  - Treatment effect 0; total variance 1;
  - Varying study sizes (zero-truncated negative binomial, \( CV=0.7 \));
  - Approximate (true) \( I \)-squared’s high: 80\% to 97\%\(^1\); moderate: 60\% to 75\%\(^2\); low: 5\% to 20\%\(^3\).

\[ \begin{align*}
\text{1} \quad \text{Equivalent} \quad \tau_{CT}^2 &= 0.25; \quad \tau_C^2 = 0.125; \quad \sigma_e^2 = 0.625. \\
\text{2} \quad \text{Equivalent} \quad \tau_{CT}^2 &= 0.125; \quad \tau_C^2 = 0.125; \quad \sigma_e^2 = 0.75. \\
\text{3} \quad \text{Equivalent} \quad \tau_{CT}^2 &= 0.025; \quad \tau_C^2 = 0.125; \quad \sigma_e^2 = 0.85. \\
\text{4} \quad \text{Equivalent} \quad \tau_{CT}^2 &= 0.0025; \quad \tau_C^2 = 0.125; \quad \sigma_e^2 = 0.8725.
\end{align*} \]
Correlation between $I^2$ one-stage and $I^2$ two-stage

- Good correlation in large samples or when $I^2$ high;
- Models failed to converge for some scenarios with very low $I^2$.

---

1 Austin (2018) The effect of number of clusters and cluster size on statistical power.
Investigation of bias in small sample settings

No clear method preferable

I-squared known to exhibit bias (upward for low I-squared, downward for high I-squared)\(^1\); no clear differences identified between two metrics.

\(^1\)The heterogeneity statistic I\(^2\) can be biased in small meta-analyses Paul T von Hippel BMC Med Res Methodol. 2015;
Points for discussion

- Quantifying treatment effect heterogeneity:
  - Caution: Low I-squared can indicate no treatment heterogeneity or insufficient evidence to make conclusive statements.
  - Caution: High I-squared can indicate clinically important treatment effect heterogeneity or very large sample sizes\(^1\).
  - Caution: I-squared can provide ball-park descriptions of magnitude of heterogeneity; best used in conjunction with a predictive interval.

- The proposed I-squared has the potential to be used in:
  - In cluster trials where treatment is crossed with cluster (to describe treatment effect heterogeneity across clusters);
  - In individually randomised trials (to describe treatment effect heterogeneity across sites).

---


\(^2\)Chen B, Benedetti A. Quantifying heterogeneity in individual participant data meta-analysis with binary outcomes Syst Rev. 2017; 6: 243
Thank you!
Poor man’s estimate of the variance of the average within cluster treatment effect

Recall, for trials where treatment is crossed with cluster:

\[ I^2 = 100 \times \frac{\tau^2}{\tau^2_{CT} + \frac{4\sigma^2_e}{S\bar{m}}} \]

The (average) within-cluster estimate of the variance of the treatment effect is estimated by:

\[ \frac{4\sigma^2_e}{S\bar{m}} \]

But....

Whilst this is correct for large sample continuous outcomes where there are no time effects, it should ideally be the average of the variance of within-cluster treatment effects. These are not a direct estimate of the modeling.
Bland Altman plot

Difference $[I^2_{\text{one-stage}} - I^2_{\text{two-stage}}]$

Average $[(I^2_{\text{one-stage}} + I^2_{\text{two-stage}})/2]$
Recall, for trials where treatment is crossed with cluster:

$$I^2 = 100\% \times \frac{\hat{\tau}_{CT}^2}{\hat{\tau}_{CT}^2 + \hat{\sigma}_e^2 \sum_j \frac{1}{m_j} \left(\frac{1}{s_j} + \frac{1}{(S-s_j)}\right)}$$

(7)

where $s_j$ denotes the number of time periods that cluster $j$ is observed under the intervention condition.