Individual participant data meta-analysis for a binary outcome: one-stage or two-stage?

Thomas Debray

Moons KGM, Abo-Zaid GMA, Koffijberg H, Riley RD

Supported by Netherlands Organisation for Scientific Research (TOP 9120.8004, 918.10.615 and 916.11.126) and the MRC Midlands Hub for Trials Methodology Research (Medical Research Council Grant ID G0800808)
Statement of Disclosure

I have no actual or potential conflict of interest in relation to this presentation
Epidemiological studies

Risk factor or predictor finding studies
- Estimation of associations
- Meta-analysis
  - Aggregate Data (AD) notoriously prone to bias
  - Individual Participant Data (IPD) more reliable!
- Approaches
  - Two-stage: reduce IPD to AD, and summarize AD
  - One-stage: synthesize the IPD from all studies in a single step

How do these methods compare, and how can they account for covariates?
Statistical Analysis of Individual Participant Data Meta-Analyses: A Comparison of Methods and Recommendations for Practice

Gavin B. Stewart¹*, Douglas G. Altman², Lisa M. Askie³, Lelia Duley⁴, Mark C. Simmonds¹, Lesley A. Stewart¹

“For these data, two-stage and one-stage approaches to analysis produce similar results.”

“Researchers considering undertaking an IPD meta-analysis should not necessarily be deterred by a perceived need for sophisticated statistical methods [...].”
Motivating Example

Diagnosis of Deep Venous Thrombosis (DVT)
- IPD from 13 studies with 10,002 patients
- Outcome: DVT presence (binary)
- Predictors: patients’ history, physical examination and results from a biomarker test
- Between-study heterogeneity

Goal: obtain pooled log odds ratios
Two-stage models

Stage 1

- For each study fit a logistic regression model with intercept $\alpha$ and slope $\beta$ (log odds ratio).
- $y_i \sim \text{Bernoulli}(p_i)$ with $\text{logit}(p_i) = \alpha + \beta x_i$

Stage 2

- Summarize estimates from stage 1
- Univariate meta-analysis: $\beta_j \sim \mathcal{N}(\mu, \tau)$
- Bivariate meta-analysis: $(\alpha_j, \beta_j) \sim \mathcal{N}(M, \Sigma)$

Estimation procedures: MLE, REML and MOM
Two-stage models

Estimation issues: zero cell counts (stage 1), correlation between random effects (stage 2)
One-stage models

Multilevel (hierarchical/mixed effects) model
- Clustering of patients within studies
Random intercept and random slope
- \( y_{ij} \sim \text{Bernoulli}(p_{ij}) \) with \( \text{logit}(p_{ij}) = \alpha_j + \beta_j x_{ij} \)
- Joint distribution for \( \alpha_j \) and \( \beta_j \) (bivariate MA)
- \( \alpha_j \) and \( \beta_j \) independently distributed (univariate MA)

Stratified intercept and random slope
- Estimate intercept for each study (rather than its distribution)
- \( \text{logit}(p_{ij}) = \sum_m \alpha_m I_{m=j} + \beta_j x_{ij} \) and \( \beta_j \sim \mathcal{N}(\mu, \tau) \)
Few/no REML procedures due to the computational difficulty in Laplace approximation; convergence issues MLE!
For each study: \( \logit(p_i) = \alpha + \beta x_i + \sum_k \theta_k z_{ik} \)

**Estimation in stage 1 sometimes problematic (zero cell counts in small studies)**
Examining Multiple Risk Factors: one-stage models

Few/no REML procedures due to the computational difficulty in Laplace approximation; convergence issues MLE!
Meta-analysis: different estimates for pooled effects, SE, heterogeneity and correlation

- Method
  (one-stage, two-stage)
- Estimation procedure
  (MLE, REML, MOM, #quadr points)
- Model specification
  (univariate, multivariate, stratified)

- One-stage models generally more reliable
- Stratified models tends to reduce SE and heterogeneity; no need to estimate correlation

Meta-analysis method should be pre-specified in study protocol