Statistical methods for individual participant data meta-analysis

Catrin Tudur Smith¹ and Richard Riley²

¹MRC North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool
²Unit of Public Health, Epidemiology & Biostatistics, University of Birmingham

Acknowledgements

• Alex Sutton, University of Leicester
• Tony Marson, University of Liverpool
• Ruwanthi Kolamunnage-Dona, University of Liverpool
• Paula Williamson, University of Liverpool

Outline

• Introduction to IPD
• Statistical methods
  – Two-stage
  – One-stage
• Examining covariates
• Bias in meta-analysis

Not Covered in detail

• How to collect IPD
• How to manage IPD
• How to check IPD

See the following examples for further details:
doi:10.1002/9780470047731.dch018

Participant Experiences

• Are you undertaking an IPDMA?
  – Yes → why IPD?
  – No → reason for coming to the workshop?
• How successful have you been at getting IPD?
• What methods did you use for doing so?
• What sort of outcomes do you have in your review?
• What are your views of benefits / disadvantages of trying to do an IPD analysis?

Aggregate Data (AD) published

Aggregate Data (AD) published

**Individual participant data (IPD)**

<table>
<thead>
<tr>
<th>Patient Number</th>
<th>Treatment</th>
<th>Survival Time (Days)</th>
<th>Status</th>
<th>Age</th>
<th>Sex</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>E</td>
<td>44</td>
<td>Dead</td>
<td>67</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>54</td>
<td>Dead</td>
<td>64</td>
<td>m</td>
<td>III</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>67</td>
<td>Alive</td>
<td>55</td>
<td>f</td>
<td>III</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>86</td>
<td>Alive</td>
<td>59</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>97</td>
<td>Alive</td>
<td>79</td>
<td>f</td>
<td>IV</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>90</td>
<td>Alive</td>
<td>59</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>90</td>
<td>Alive</td>
<td>59</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>90</td>
<td>Alive</td>
<td>59</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>90</td>
<td>Alive</td>
<td>59</td>
<td>m</td>
<td>IV</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>23</td>
<td>Dead</td>
<td>53</td>
<td>m</td>
<td>IV</td>
</tr>
</tbody>
</table>

**Aggregate Data (AD) requested**

- Be aware there is another option
- Specific summary tables / statistics beyond those in paper may be sufficient / more desirable than relying on published result to carry-out the analysis required
- Desirable if investigators are unwilling to hand over their dataset but would supply further information

---

**Why IPD?**

M-A of IPD is considered the gold standard approach to analysis

- Used relatively infrequently
- But, becoming more common (Simmonds et al 2005) – 79 IPD reviews on the IPDMWG site

General disadvantages

- Time consuming and costly
- May not be able to obtain all IPD - retrieval bias

---

**Why IPD?**

M-A of IPD is considered the gold standard approach to analysis

- Used relatively infrequently
- But, becoming more common (Simmonds et al 2005) – 79 IPD reviews on the IPDMWG site

General disadvantages

- Time consuming and costly
- May not be able to obtain all IPD - retrieval bias
Examples of when IPD?

- High patient exclusion rate
  - IPD meta-analysis in soft tissue sarcoma (1997), authors recovered data on 99% of the 344 patients that had been excluded from individual trial analyses.
  - Without additional patients HR=0.90 (p=0.05); including additional patients HR=0.85 (p<0.05)

- Time-to-event data
  - Epilepsy example, time to 12 month remission and time to treatment failure recognised as outcomes of clinical importance
  - Most trials reported different outcomes (50% reduction in seizures) or different definitions (time from specific dose level rather than time from randomisation), or
  - Trials did not report sufficient summary data to allow HR to be estimated reliably

Examples of when IPD?

- When interactions with treatment are important
  - Interaction between epilepsy type and treatment
  - Heterogeneity across studies in AD meta-analysis of cervical cerclage

- To investigate longer-term outcomes
  - such as for chronic diseases where events take place over a long period of time

- Meta-analysis of prognostic factors studies
  - Use a (small) consistent set of adjustment factors across studies
  - Use a consistent cutpoint across studies, or produce continuous marker results.
  - Assess the benefits of using combinations of markers

IPD vs AD

- IPD and AD meta-analysis can be equivalent
  - if data are equivalent
  - if treatment effect measure are equivalent

IPD vs AD

- Discrepancies usually arise because IPD data sets include different data to AD
  - IPD may reinstate patients originally excluded
  - IPD may include additional follow-up data
  - IPD may use more appropriate effect measure

IPD vs AD

<table>
<thead>
<tr>
<th></th>
<th>AD  (11 trials 1911 patients)</th>
<th>IPD  (13 trials 2103 patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR 0.65 (95% CI)</td>
<td>0.53 to 0.83</td>
<td>0.83 (95% CI)</td>
</tr>
</tbody>
</table>

**IPD vs AD**

"...the balance of gains and losses of the approach will vary according to the disease, treatment, and therapeutic questions explored"  Stewart and Tierney 2002

- Empirical evidence - precision and size of effect varies compared to AD but no systematic pattern
- Further empirical evidence is needed:
  - Individual patient data meta-analyses compared with meta-analyses based on aggregate data. Clarke MJ, Stewart L, Tierney J, Williamson PR
  - Protocol for methodology review – Cochrane Library

SEE POSTER HERE!

**Meta-Analysis of IPD**

- Decisions for analysis with IPD
  - Two-stage or One-stage
  - Fixed or Random Treatment Effects
  - Which software?
  - Prognostic Factors and Effect Modifiers

**Meta-Analysis of IPD**

- Common approaches to M-A of IPD include
  - **Two stage**: create summary statistics out of IPD (stage 1) and combine using standard meta-analysis method (stage 2)
    - either fixed effect or random effects approach
  - **One stage**: combine all patient data from all studies in one single model taking into account the clustering of patients within study
    - either fixed effect or random effects approach

**Common practice**

- Simmonds et al (2005), n=44, 1999-2001
  - 65% with <=10 trials
  - two-stage methods most common
  - poor reporting
- Pignon et al (2007), lung cancer, n=9, - 2006
  - two-stage methods most common
- Kolamunnage-Dona (2008), n=79 (62 with data on number of trials), IPDMWG
  - median 10 trials, range 2-63
  - two-stage methods most common

**Two-stage : time to event data**

**STAGE 1**: For each trial separately, reduce the IPD to Aggregate Data

(i) Fit a separate Cox proportional hazards model (IPD) to each trial e.g proc phreg (by trial) in SAS

(ii) Obtain an estimate of log hazard ratio and its standard error for each trial

> THIS IS AGGREGATE DATA

**STAGE 2**: Combine the AD using standard meta-analysis methods

(i) Enter data into meta-analysis software e.g. Revman
(ii) Undertake meta-analysis using generic inverse variance method (either fixed effect or random effects)
    - Calculate I-squared as usual
(iii) Could extend to adjust for covariates within trials
(iv) Extend to meta-regression if required
**Two-stage : time to event data**

- **ALTERNATIVE:** Stratified log-rank analysis (Early Breast Cancer Trialists’ Collaborative Group)

**STAGE 1:** Undertake a log-rank analysis for each trial to obtain estimates of

The log-rank statistic $U_j$ and its variance $V_j$

**STAGE 2:** Combine over all trials using

$$\hat{\beta} = \frac{\sum_j U_j}{\sum_j V_j} \quad \text{with} \quad SE(\hat{\beta}) = \sqrt{\sum_j V_j}$$

**Two Stage: continuous or binary data**

**Stage 1:** Fit a separate model for all patients in each trial and extract estimates of treatment effect and standard error

- **Continuous Data**
  - Normal linear regression model – estimate of difference in means and standard error

- **Binary Data**
  - Logistic regression model – estimate of log odds ratio and standard error

**Stage 2:** Pool across trials using standard meta-analysis methods

**Software for Two Stage Approach**

- **Using Revman (free)**
  
  **Stage 1:**
  
  Use a standard Stats Package to obtain estimates of treatment effect and SE eg SAS (proc phreg), R (coxph), STATA (stcox)

  **Stage 2:**
  
  Input data using Generic Inverse Variance Method in Revman
  
  Note: ‘O - E and Variance’ option in Revman fits ‘Peto’ method if logrank ‘O - E’ and ‘Y’ statistics have been obtained.

  This is a FE analysis – no equivalent RE analysis is available in RevMan.

**Software for Two Stage Approach**

- **Using SCHARP - Survival Curve and HAZard Ratio Program**
  - an interactive SAS-based application
  - Analyses and plots IPD meta-analyses
  - Uses two-stage approach
  - Analysis of time-to-event, dichotomous and continuous outcomes with choice of measures
  - Fixed effect and random effects models
  - Version 4 available but still need SAS and SCHARP may still have bugs
  - Free from MRC Clinical Trials Unit, London

**Two-stage Approach**

- **Benefits:**
  - Straightforward
  - Accommodate Fixed and Random Effects (using DerSimonian-Laird method in usual way)
  - ‘Standard’ meta-analysis interpretation: Forest Plot and Heterogeneity statistics output
  - Can easily incorporate both IPD and AD estimates within the same meta-analysis

- **Limitations**
  - Cumbersome with many trials
  - Limited - cannot fully investigate patient level effect modifiers

**One-stage regression models for time-to-event data**

- Investigating heterogeneity in an individual patient data meta-analysis of time to outcome events

  Caitin Tsuchi Smith1,*, Paula R. Williams1 and Anthony G. Manzo1

  1Centre for Clinical Trials and Health Evaluation, University of Liverpool, Liverpool L69 7AE, U.K.
  2Department of Epidemiological Sciences, University of Liverpool, Liverpool, U.K.

  **SUMMARY**

  Different forms of studies in terms of design, balance and methodology, clinical procedures, and patient characteristics, as factors that can contribute to variability in the treatment effect. Short-term follow-up studies in different settings may provide a wide range of evidence that can be used to assess the consistency between treatment effect and outcome with the aim of evaluating the variability in terms of clinical and methodological differences. Such an investigation can be conducted using aggregated data or via IPD meta-analysis. As a result, heterogeneity analysis can provide a better understanding of potential sources of heterogeneity and modify a further analysis of future cases in event selection or meta-analysis.
One-stage regression models for time-to-event data

Data from each patient in each trial are included in a single model eg Cox model

Fixed Effect (SFE/FE)
Cox model stratified by trial with fixed treatment effect
Assume proportional hazards within trials, but not across trials
Treatment heterogeneity assessed via trial-specific effects
No allowance is made for residual heterogeneity

Random Effects (SFE/RE)
Cox model stratified by trial with random treatment effects
Assume proportional hazards within trials, but not across trials
Treatment heterogeneity assessed via heterogeneity parameter
Allowance is made for residual heterogeneity

Software for one-stage regression models for time-to-event data

- Cox models with fixed effects fitted using standard statistical software eg proc phreg in SAS, coxph in R, stcox in STATA
- Random effects models
  - SAS IML: approach outlined by Yamaguchi (1999), adapted by Tudur Smith (2005) (fixed trial, stratified or random trial)
  - R coxme: still under development?
- Abrams note that you can re-formulate Cox model as Poisson regression model (Whitehead, 1980; Lindsey, 1995)
  - Relatively easy to specify random effects
  - Implemented in R using lmer function
- Other estimation methods for random effects models - Cortinas Abrahantes et al (2007) compared 4 methods but were not able to make any clear recommendation

Other alternatives described by Tudur Smith et al 2005.

Comparison of methods for time-to-event data

- Comparison of five alternative 'one-stage' Cox models
  1
- Stratified Cox model appropriate for most situations
  - More appropriate to assume different baseline hazard in each trial rather than a common baseline across all trials
  - Trial effects as dummy variables compares patients across trials
  - Trial effects as random effects may not be reasonable?
  - Computationally more efficient to fit for random treatment effects
- NOTE: if many trials included may produce unstable estimates,
  - Efficiency gains for random trial effects greatest for moderate to large numbers of very small groups (of sizes two or three)
  1 Tudur Smith and Williamson (Stat Med 2005)

Comparison of methods for time-to-event data

- How does the stratified Cox model (one-stage) compare with two stage approaches?
  - Stratified Cox versus Stratified logrank versus IV Cox
  - All assuming fixed treatment effects
  - Simulation study: 5 trials, 100 patients in each group, 1000 simulations

Comparison of methods for time-to-event data

- How does the stratified Cox model (one-stage) compare with two stage approaches?
  - Stratified Cox versus Stratified logrank versus IV Cox
  - All assuming fixed treatment effects
  - Simulation study: 5 trials, 100 patients in each group, 1000 simulations

Mean of log hazard ratios over 1000 simulations

$\tau^2 = 0$

<table>
<thead>
<tr>
<th>True log hazard ratio</th>
<th>$\tau^2 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

TRUE IV pooled strat cox strat logrank

Comparison of methods for time-to-event data

- How does the stratified Cox model (one-stage) compare with two stage approaches?
  - Stratified Cox versus Stratified logrank versus IV Cox
  - All assuming fixed treatment effects
  - Simulation study: 5 trials, 100 patients in each group, 1000 simulations

No heterogeneity
  - all methods perform well for small effects as expected theoretically
  - stratified logrank displays bias and poor coverage for larger effects

Increasing heterogeneity
  - Coverage decreases quite dramatically, bias increases
An example from epilepsy

- 5 RCTs with 1225 patients comparing CBZ and SV monotherapy (3 other eligible trials did not collect seizure data)
- Time to 12-month remission - not reported sufficiently in any of the trials
- Clinically important covariates: age, gender, epilepsy type, log(no. seizures), time from first ever seizure

Epilepsy: 12-month remission

- Model without covariates

  Heterogeneity: p=0.02, I² = 66%
  SFE/FE: log HR = 0.132 (0.073)
  SFE/RE: log HR = 0.098 (0.125), τ² = 0.0484 (0.055)

  Adjustment for: age, epilepsy type, log(seizures), age*treatment

  SFE/FE: log HR = -0.162 (0.129)
  SFE/RE: log HR = -0.163 (0.139), τ² = 0.006 (0.027)

  τ² reduced by 88%

One-stage models for other data

1. **Continuous Outcomes** (see Higgins JPT. et al. Stat Med 2001)
   - Fixed effect - standard ANOVA model
   - Random effects - SAS PROC MIXED, MLwiN, STATA xtmixed, winBUGS

   - Generally based on logistic regression models
   - Fixed effect models - standard stats software eg SAS, R, STATA
   - Random effect models – MLwiN, STATA glamm, winBUGS

   - Based on proportional odds models

Summary

- Decisions for analysis include:
  - Two-stage or One-stage?
    - Two-stage may be more straightforward for simple cases
    - Useful if IPD not available for all trials
    - But limited: One-stage more flexible
  - Fixed or Random Treatment Effects?
    - Usual consideration for meta-analysis
    - Accommodated by 1-stage or 2 stage
    - Random effects can be more complex in 1-stage framework
  - Which software?
    - Depends on data and model
    - Random effects survival analysis using SAS IML but not very efficient
    - Multilevel model software for other outcomes
  - Prognostic Factors and Effect Modifiers
    - Use one-stage approach,...
Using IPD meta-analysis to examine treatment-covariate interactions

Richard D. Riley
Dept of Epidemiology, Public Health & Biostatistics, & School of Mathematics
University of Birmingham
e-mail: r.d.riley@bham.ac.uk

Many thanks to:
Paul Lambert

Outline

Aim:
to show a range of examples where IPD meta-analysis helps examine treatment-covariate interactions

• One-step and two-step approaches

• Differences to a meta-regression of summary data

• Understanding threat of ecological bias

Part 1:
Rationale and Methods

Stratified medicine
• Increasing interest in personalised or stratified medicine

• We want to tailor treatment to individuals, or to groups of similar individuals

• To do this, we need to identify individual-level factors (covariates) that modify treatment response

• Essentially, what factors cause some patients to respond better to treatment than others?

Stratified medicine
• Statistically, this means we want to examine & estimate so-called treatment-covariate interactions
  - i.e. quantify how particular covariates interact with treatment effect

• Also known as subgroup effects & effect modifiers

• Individual studies usually have low power to detect them, as they are powered on the overall treatment effect (the average across all individuals)

• By combining studies, meta-analysis thus offers an opportunity to increase power to detect true treatment-covariate interactions

Stratified medicine
• For commissioners of healthcare
  - stratified medicine offers the potential to maximise treatment related benefit and reduce treatment related harm.

• For developers of new interventions
  - stratification may offer the opportunity to rescue a treatment which fails to show overall benefit in unselected patients, but that might have worthwhile benefit in an identifiable subgroup.
Example: Estrogen receptor in breast cancer

- Tamoxifen is only given to patients who are ER positive, as an IPD meta-analysis found ...

**ER +ve**

**ER unknown**

**ER -ve**

All patients

### IPD Methods: A two-stage approach

- So how do we estimate treatment-covariate interactions in an IPD meta-analysis?

  - Let us consider an example with continuous data

- There are 10 trials in hypertension (high blood pressure) and we are interested in the treatment effect on systolic blood pressure (SBP)

- Also interested in treatment-covariate interactions

### Continuous data IPD

<table>
<thead>
<tr>
<th>Study</th>
<th>Patient</th>
<th>SBP initial</th>
<th>SBP final</th>
<th>treat</th>
<th>placebo</th>
<th>age</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>190</td>
<td>185</td>
<td>1</td>
<td>0</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>175</td>
<td>172</td>
<td>1</td>
<td>0</td>
<td>69</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>184</td>
<td>185</td>
<td>0</td>
<td>1</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>192</td>
<td>182</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>201</td>
<td>199</td>
<td>1</td>
<td>0</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>169</td>
<td>154</td>
<td>1</td>
<td>0</td>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>171</td>
<td>170</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>179</td>
<td>168</td>
<td>0</td>
<td>1</td>
<td>67</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>197</td>
<td>167</td>
<td>1</td>
<td>0</td>
<td>83</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>189</td>
<td>171</td>
<td>1</td>
<td>0</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>184</td>
<td>188</td>
<td>0</td>
<td>1</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>168</td>
<td>161</td>
<td>0</td>
<td>1</td>
<td>61</td>
<td>0</td>
</tr>
</tbody>
</table>

Can see that treatment effects, baseline factors and prognostic variables are available per individual (note data are truncated for each study, as actually hundreds of patients)

### IPD Methods: A two-stage approach

(a) Overall treatment effect

Step 1: Estimate the treatment effect and its variance in each IPD study using an appropriate method, such as analysis of covariance

Step 2: Take the effect estimates for each study, and combine them in a usual random-effects meta-analysis

- Gives a pooled treatment effect (across all individuals) of -10.16 (95% CI: -12.27 to -8.06).

- So hypertension treatment is significantly effective in reducing systolic blood pressure by, on average, 10.16 mm Hg more than control.

(b) Effect of sex on the treatment effect

- Let 1 = males, and 0 = females

Step 1: Estimate the interaction between sex covariate and treatment effect, and its variance, in each IPD study separately using an appropriate method, such as analysis of covariance

Step 2: Take the interaction estimates for each study, and combine them in a usual fixed-effect of random-effects meta-analysis

- Treatment effect for females

- Treatment effect for males

STEP 1: \( SBP_{ij} = \beta_0 + \beta_1 SBP_{ij} + \beta_2 sex_{ij} + \beta_3 treat_{ij} + \gamma_i (sex_{ij} \times treat_{ij}) + \epsilon_{ij} \sim N(0, \sigma^2) \)

Control effect Treatment-sex interaction Residual error

\( \gamma_i = \text{Change in treatment effect for males compared to females} \)
(b) Effect of sex on the treatment effect
- sex = 1 for males, and 0 for females;
- treat = 1 for treatment, and 0 for control
Let i = study, and j = patient

STEP 2: e.g. fixed-effect meta-analysis
\[ \hat{\gamma}_i = \gamma + \epsilon_i \]
\[ \epsilon_i \sim N(0, \text{var}(\hat{\gamma})) \]

Best estimate, across all studies, of the difference in treatment effect for males compared to females

IPD Methods: A two-stage approach

- Alternatively, one can undertake a one-stage approach
- The IPD from all trials are analysed simultaneously
- Clustering of patients within trials accounted for
- Quicker and obtain multiple summary estimates together
- Obtains very similar estimates to two-stage approach

However ...
- Including interactions requires careful separation of within-study and across-study relationships (Riley et al.)
- Essentially, you can explain both within-study variability and between-study variability ... so need to separate these things out to avoid ecological bias (more later)

If IPD are not available ...
- Hope that study authors report the treatment-covariate interactions .... unfortunately rare
- Often one can only do a meta-regression
i.e. regess the study (i) treatment effect estimates (\( \theta_i \)) against average patient-level covariates
\[ \hat{\theta}_i = \alpha_i + \gamma(\text{proportion male}) + u_i + \epsilon_i \]
\[ \epsilon_i \sim N(0, V(\hat{\theta})) \]

Called the ‘across-study interaction’. Tells us how much the average treatment effect differs in a study with only males compared to a study with only females
- crucially, this is different to the ‘within-study interaction’ obtained by analysing the IPD (Riley et al., 2008)
Within-study versus between-study interactions

Within-study interaction (from IPD)
- Effect of individual covariates on treatment effectiveness
- Results tailored to individual patient
- e.g. the treatment effect for males compared to females is …
- Explains within-study variability (residual error)

Across-study interaction
- How mean patient-level covariate in a study is associated with the mean treatment effect
- Results relate to the study-level (population)
- e.g. In a population with a proportion of 70% males, the underlying mean treatment effect is …
- Explains between-study variability

Example 1: Application to hypertension data
- How does being male modify treatment effect on SBP?
  - Within-study effect
    - $\gamma_w = 0.77 \ (-0.5$ to $2.05)$
    - if for females the treatment reduces SBP by 20 mmHg more than placebo
    - then for males the treatment reduces SBP by 19.23 mmHg more than placebo
    - non-significant

  - Across-study effect
    - $\gamma_h = 15.02 \ (8.98$ to $21.1)$
    - if female studies have an underlying treatment effect that reduces SBP by 20 mmHg
    - then male studies have an underlying treatment effect that reduces SBP by 4.98 mmHg
    - significant

Example 1: graphical illustration
- Within-study effects meaningful to individual patient
- But usually not obtainable if IPD not available
- Across-study effects meaningful at the population level
- Available when mean covariate is available for each study
- Simulation studies show that in ideal conditions across-study interactions will reflect within-study interactions (‘unbiased’)
- But across-study effects have low power, & prone to ecological bias & confounding across studies: Interpret with caution!
- e.g. studies with high proportion male may also have a higher dose of treatment; thus trend in treatment effect due to dose of drug and not proportion male

Example 1: Application to hypertension data
- How does being male modify treatment effect on SBP?
  - Within-study effect
  - $\gamma_w = 0.77 \ (-0.5$ to $2.05)$
  - if for females the treatment reduces SBP by 20 mmHg more than placebo
  - then for males the treatment reduces SBP by 19.23 mmHg more than placebo
  - non-significant

  - Across-study effect
    - $\gamma_h = 15.02 \ (8.98$ to $21.1)$
    - if female studies have an underlying treatment effect that reduces SBP by 20 mmHg
    - then male studies have an underlying treatment effect that reduces SBP by 4.98 mmHg
    - significant

VERY DIFFERENT CONCLUSIONS, DUE TO ECOLOGICAL BIAS / CONFOUNDING

Example 1: graphical illustration
- Within-trial interactions (from IPD studies)
- Across-trial interaction (from meta-regression)
Example 2: Increased power to detect true covariate interactions (Lambert et al., 2004)

- 1000 meta-analyses simulated, each with 5 trials and treatment effective for high risk patients but ineffective for low risk patients.
- Each meta-analysis analysed first using IPD, and then using meta-regression: treatment-covariate interactions estimated in both cases
- The % of 1000 meta-analyses that detect this true treatment-covariate interaction with statistical significance gives the power
- The % is usually far higher when using within-study interactions from IPD than when using interactions from meta-regression
- Only when there is large between-study variation in the mean covariate value does the power of meta-regression appear adequate … but even then ecological bias and confounding may occur.

Example 2: graphical illustration

IPD approach has a power of 90.8%
Meta-regression approach has a power of 10.8%

Example 3: Effect of elevated panel reactive antibodies on the effectiveness of anti-lymphocyte antibody induction (Berlin et al., 2022)

- Meta-analysis of five randomised trials of anti-lymphocyte antibody induction therapy for renal transplant patients
- Interested in the difference in treatment effect between patients with elevated antibodies compared to non-elevated
- A meta-regression is used to examine the across-trials interaction:
  - estimated difference in log odds of treatment failure between a trial with only elevated patients compared to a trial with only non-elevated patients = -0.01 (p = 0.68)

- Did the authors need IPD to obtain this result?
- What would you conclude from this about whether treatment effect is different for elevated and non-elevated patients?

Example 3: Effect of elevated panel reactive antibodies on the effectiveness of anti-lymphocyte antibody induction (Berlin et al., 2022)

- The reviewers also estimate the pooled within-study interaction estimated difference between elevated and non-elevated patients in the log-odds of treatment failure = - 1.33 (p = 0.01)

- Did the authors need IPD to obtain this result?
- Suggest potential reasons why there is a substantial difference between within-study & across-study interactions?
- Is there a genuine difference in treatment effect between elevated and non-elevated patients?

THANK YOU!  r.d.riley@bham.ac.uk

Useful References


Are IPD meta-analyses biased?

Richard D. Riley
Dept of Epidemiology, Public Health & Biostatistics, & School of Mathematics
University of Birmingham
e-mail: r.d.riley@bham.ac.uk

Many thanks to: Alex Sutton and Ikhlaaq Ahmed
Outline

Aim:
to discuss potential biases that may affect IPD meta-analyses and to provide examples

- Publication bias
- Selection bias
- Unavailable data

Part 1:
Possible biases

Publication & related biases
- Publication bias occurs when studies with statistically significant or clinically favourable results are more likely to be published than studies with non-significant or unfavourable results.
- Other related biases exist such as:
  - time-lag bias
  - selective outcome reporting
  - language bias
  - duplication bias, etc
- Leads to meta-analyses which:
  - synthesise an incomplete set of the evidence
  - produce summary results potentially biased toward favourable treatment effects.

Publication bias: evidence for concern
- Burdett et al. found that in 9 of 11 IPD reviews, the meta-analysis result was closer to the null when ‘gray literature’ studies were included (e.g. unpublished trials, conference abstracts, etc)
- Ahmed et al. (submitted) examined 31 IPD meta-analyses of trials published between 2007 and 2009
  - Only 9 of the 31 articles included ‘gray literature’ IPD in their primary meta-analysis.
  - Thus majority (65%) do not include IPD from ‘gray literature’
- This emphasises why obtaining IPD does not automatically remove the potential for publication related biases in meta-analysis.
  - Despite this, only 10 of the 31 articles discussed or examined statistically the threat of publication bias in their IPD meta-analysis

Is IPD meta-analysis really the gold-standard?
- The phrase ‘gold-standard’ is often used in articles to describe the IPD meta-analysis approach
- We have already discussed many reasons why IPD is preferable over a traditional meta-analysis of aggregate data from publications
- Yet there has been little consideration of how potential biases may impact upon IPD meta-analyses
- For example, biases may act in:
  - the identification of relevant studies,
  - the decision about which studies to seek IPD from,
  - the amount of IPD obtained from studies,
  - the type of studies that agree to provide their IPD
- IPD allows trial results to be derived directly and independent to study reporting
  - it thus has potential to reduce publication and related biases
  - especially if IPD are obtained for unpublished trials
- Yet, all these bias problems hide pertinent trials and their results
- Thus – just as in a standard systematic review and meta-analysis – they may cause IPD researchers to miss relevant but non-significant trials
- Hence, IPD from relevant trials may not be sought
**Data availability bias**
- IPD may not be available for all studies
- If unavailability of IPD is related to the study results, this may cause bias
- The impact of availability bias is hard to predict
  e.g. (1): Researchers of studies with non-significant results may be more likely to have destroyed or lost their IPD
  - bias IPD meta-analyses toward a favourable treatment effect
  e.g. (2): Researchers of studies with favourable findings may not provide their IPD because they want to utilise it further, for subgroup effects or an extended follow-up
  - bias IPD meta-analyses toward a lower treatment effect

**Data availability bias: evidence for concern**
- Reasons for unavailable IPD include:
  - trial data being lost or destroyed;
  - study authors not being contactable, unwilling to collaborate or unable to send their data
- When IPD are unavailable for some trials, the IPD approach may not be better than a meta-analysis of aggregated data from all trials.
- Investigate the potential impact of non-IPD trials on IPD meta-analysis conclusions, wherever possible.
  e.g. Vale et al. obtain aggregate results for three of their ten missing trials, and 'incorporating them into the meta-analysis did not materially change the results.'
- Statistical approaches which synthesise both IPD and aggregate data are potentially valuable here (Riley et al., 2008)

**Selection bias**
- Selection bias occurs if reviewers only seek IPD from a subset of existing studies, and this subset does not reflect the evidence-base.
- This is a particular concern when:
  - relevant studies are not identified by a systematic review but rather through contacts or friends in the field
  - when selection takes place with knowledge of study results
- The impact of selection bias is hard to predict
  - it may (directly or indirectly) be affected by the selectors’ knowledge of the field, their research contacts & collaborations, and their opinion about the research question of interest.
- It is less of a concern for prospective IPD meta-analysis: as study results are unknown at the time of study recruitment

**Selection bias: evidence for concern**
In the survey by Ahmed et al. (submitted)
- 22 of the 31 articles performed a systematic review to identify all relevant trials, from which IPD was then requested.
- In the other 9 articles selection bias is a potential concern, as identification of relevant studies was either not stated or based on a more selective, non-systematic approach.
  For example, Papakostas et al. state: ‘although we included all eligible studies sponsored by GlaxoSmithKline regardless of whether they have been published or not, it is possible that studies sponsored by other sources have been conducted but have not been yet published or presented at major scientific meetings.’
- Do not automatically view an IPD meta-analysis as ‘gold standard’ without due thought as to how IPD studies were chosen.

**Part 2: Illustrated examples**
**Example 1: Greb et al.**

- Review whether high-dose chemotherapy with autologous stem cell transplantation as part of first-line treatment improves survival in adults with aggressive non-Hodgkin lymphoma.
- 15 randomised trials comparing high-dose versus conventional chemotherapy were identified by a systematic review.
- IPD were sought from all 15 trials, so selection bias is not a concern.
- However, publication and availability biases are a threat, as:
  - 4 of 15 trials were available.
  - IPD was unavailable for 5 of them (33%).
- Greb et al. examine both these issues; now extend their work …

---

**Example 2: De Luca et al.**

- Review the benefits of early versus late use of Gp IIB-IIIa inhibitors in patients undergoing primary angioplasty for ST-segment elevation myocardial infarction.
- A primary angiographic endpoint was whether patients achieved a preprocedural Thrombolysis in Myocardial Infarction (TIMI) grade 3 flow distal embolisation.
- A systematic review identified 14 relevant trials and IPD was sought from them all, so selection bias is not a concern.
- However, availability and publication biases are a threat, as:
  - IPD was unavailable for 3 trials (23%)
  - all 11 trials providing IPD were fully published.
- De Luca et al. did not investigate these biases, so let’s do it …

---

**Greb et al: Data availability bias?**

- Impact of including aggregate results from 4 of the 5 non-IPD studies …

---

**Greb et al: Publication bias?**

- Is there any evidence of funnel plot asymmetry (a signal for potential publication bias)?
**De Luca et al: Data availability bias?**

• Impact of including aggregate results from 2 of the 3 non-IPD studies

**De Luca et al: Publication bias?**

• Is there any evidence of funnel plot asymmetry (a signal for potential publication bias)?

---

**Box 4: Suggested information to report from an IPD meta-analysis, to supplement those reporting guidelines elsewhere**

- Articles of applied IPD meta-analyses should report:
  - if there was a protocol for the IPD project, and where it can be found
  - if ethics approval was necessary and (if appropriate) granted
  - why the IPD approach was initiated originally
  - the process used to identify relevant studies for the IPD meta-analysis
  - how authors of relevant studies were approached for IPD
  - how many studies (or collaborating groups) were approached for IPD, and the proportion that provided IPD
  - whether those studies that provided IPD gave all their IPD or only a proportion; if the latter, then describe what information was omitted and why
  - whether the IPD results for each study were comparable to published results, and if not why not (e.g. IPD contained updated or modified information)
  - the number of patients within each of the IPD studies and, if appropriate, the number of events

---

**Many of the issues discussed today are in …**

*Cite this as: BMJ 2010;340:c221*

**Research Methods & Reporting**

Meta-analysis of individual participant data: rationale, conduct, and reporting

Richard D Riley, senior lecturer in medical statistics; Paul C Lambert, senior lecturer in medical statistics; Ghalie Abo-Zaid, postgraduate student

1 Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham B15 2TJ. 2 Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester LE1 7RH. 3 School of Mathematical Sciences, University of Birmingham, Birmingham B15 2TJ.

Correspondence to: R D Riley (r.d.riley@umans.ac.uk)

The use of individual participant data instead of aggregate data in meta-analyses has many potential advantages, both statistically and clinically. Richard D Riley and colleagues describe the rationale for an individual participant data meta-analysis and outline how to conduct this type of study.
**Useful References**

- Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25(1):76-97.

THANK YOU!  r.d.riley@bham.ac.uk