Individual participant data meta-analysis

Catrin Tudur Smith
cat1@liv.ac.uk

MRC North West Hub for Trials Methodology Research,
Department of Biostatistics,
University of Liverpool

Acknowledgements

- Thomas Debray, UMC Utrecht
- Richard Riley, Keele University
- Tony Marson, University of Liverpool
- Paula Williamson, University of Liverpool

Outline

- What is IPD?
- Why IPD?
- How to get and process IPD
- How to analyse IPD

  (i) treatment effect

  (ii) treatment-covariate interaction
- Further issues
- Practical session (using R)

Aggregate Data (AD) published

Table 4. Summary of Efficacy Results: Overall Survival and Progression-Free Survival

	No. of Patients	% GemCis	% Gem	HR	Log-Rank *P*
Median OS					
All patients	190	7.5	6.0	0.80	.15
Locally advanced	39	10.3	10.4	0.68	.29
Metastatic	151	7.2	4.7	0.82	.23
KPS 70%-80%	76	4.9	4.8	1.13	.64
KPS 90%-100%	84	10.7	6.9	0.62	.051*
6-month survival		59.0	50.5		.45
12-month survival		25.3	24.7		.21
Median PFS					
All patients	190	5.3	3.1	0.75	.053
Locally advanced	39	8.6	3.2	0.30	.0053
Metastatic	151	4.2	3.1	0.84	.31
KPS 70%-80%	76	2.8	2.9	0.91	.69
KPS 90%-100%	84	7.7	2.8	0.54	.013†

Abbreviations: GemCis, gemcitabine plus cisplatin; Gem, gemcitabine alone; HR, hazard ratio; OS, overall survival; KPS, Karnofsky performance status; PFS, progression-free survival.
*Peto-Wilcoxon-Test $P = .0079$.
†Peto-Wilcoxon-Test $P = .0020$.

Journal of clinical oncology 2006, *24:3946-3952*.

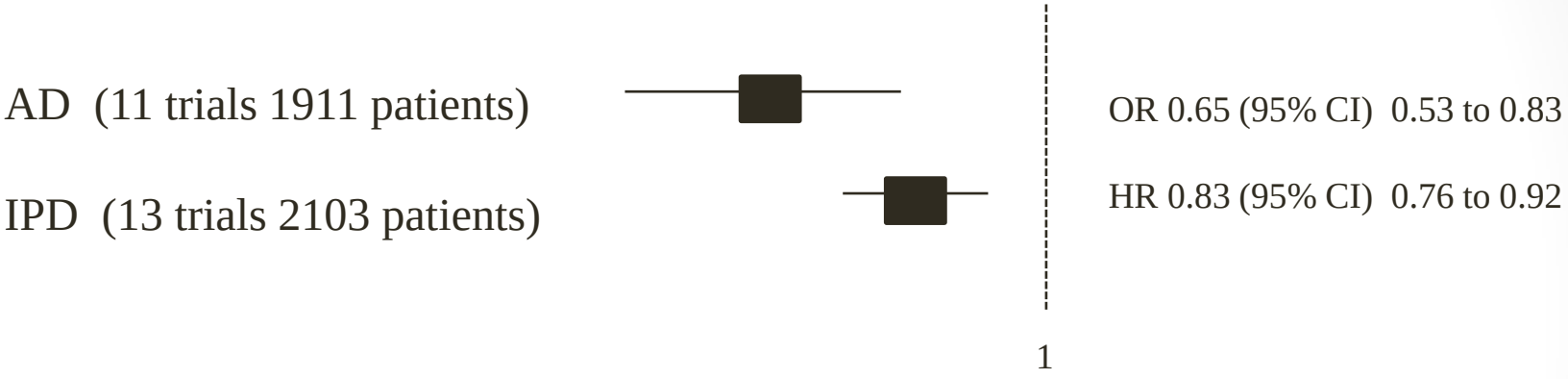
Individual participant data (IPD)

Patient Number	Treatment	Survival Time (Days)	Status	Age	Sex	Stage
1	E	44	Dead	67	m	IV
2	E	54	Dead	64	m	III
3	E	67	Alive	55	f	III
4	C	43	Dead	79	f	IV
5	C	70	Alive	62	m	IV
6	E	88	Dead	60	f	IV
7	C	99	Alive	57	m	III
8	C	45	Dead	66	m	III
9	E	90	Alive	59	f	III
10	C	23	Dead	53	m	IV

IPD vs AD

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



Pignon JP and Arriagada R. Lancet 1993.

IPD vs AD

- 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

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

Why IPD ?

Reinstate patients into the analysis who were originally excluded

- Tierney and Stewart (2005) IPD meta-analysis in soft tissue sarcoma
- 99% of the 344 patients that had been excluded from published individual trial analyses were recovered

Meta-analysis with exclusions:	HR=0.85 (p=0.06)
Meta-analysis reinstating all exclusions:	HR=0.90 (p=0.16)

Why IPD ?

Overcome outcome reporting bias (ORB)

- Definition: Selection of a subset of the original recorded outcomes, on the basis of the results, for inclusion in publication

The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews

Jamie J Kirkham,[1] Kerry M Dwan,[1] Douglas G Altman,[2] Carrol Gamble,[1] Susanna Dodd,[1] Rebecca Smyth,[3] Paula R Williamson[1]

- ORB suspected in at least one trial in 34% of 283 Cochrane reviews
- 42 significant meta-analyses
  - 8 (19%) would not have remained significant
  - 11 (26%) would have overestimated the treatment effect by > 20%

BMJ (2010); **340**:c356

Why IPD ?

Detailed exploration of participant level covariates' influence on treatment effect

- Meta-analysis of 5 RCTs of anti-lymphocyte antibody induction therapy vs control for renal transplant patients **(Berlin et al., 2002)**
- Difference in treatment effect between patients with elevated antibodies compared to non-elevated?

  - Aggregate Data to estimate across-trials interaction:
    estimated difference in log odds ratio between elevated and non-elevated patients = -0.01 (p = 0.68)

  - IPD to estimate the pooled within-study interaction:
    estimated difference in log odds ratio between elevated and non-elevated patients = - 1.33 (p = 0.01)

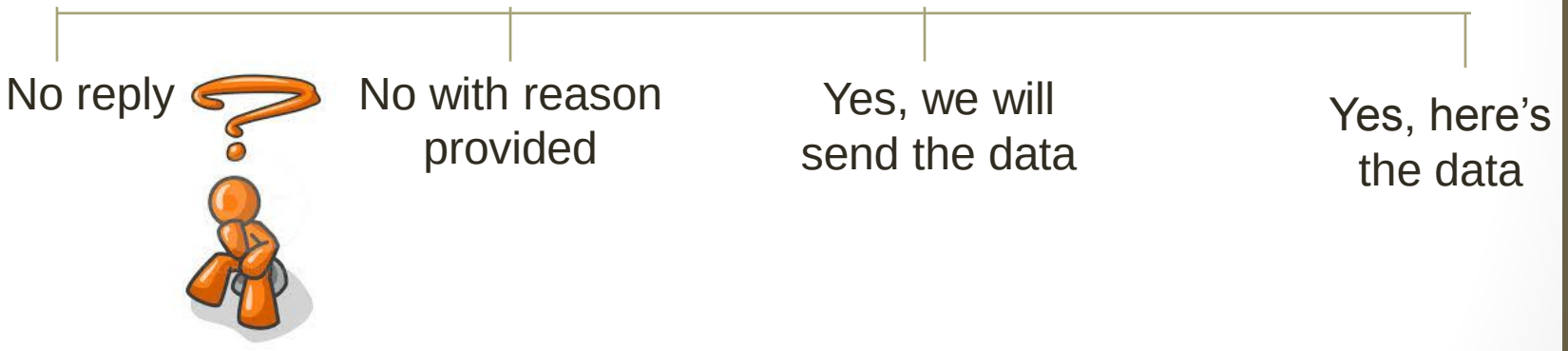
Why IPD?

Data checking & standardisation of analysis	• Outcome definition can be standardised across trials
More complete analysis	• Include follow-up beyond initial publication • Reinstate patients into the analysis who were originally excluded • May be able to overcome outcome reporting bias
Detailed exploration of participant level covariates influence on treatment effect	• Maximum information using patient as unit of analysis - more power to identify clinically moderate interaction • Direct interpretation for individual patient • No reporting bias of subgroup analyses • No ecologic bias
More thorough analysis of time-to-event data	• Check model assumptions eg proportional hazards • More accurate (if published AD restricted)

But the IPD approach will be more resource intensive!

How to get IPD

- Eligible trials identified by search as in an AD review
- Identify contact author eg email address published in journal
- Response to request can vary



- Variation in data format and supporting material

How to get IPD

- Initiatives to encourage data sharing and clinical trial transparency

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

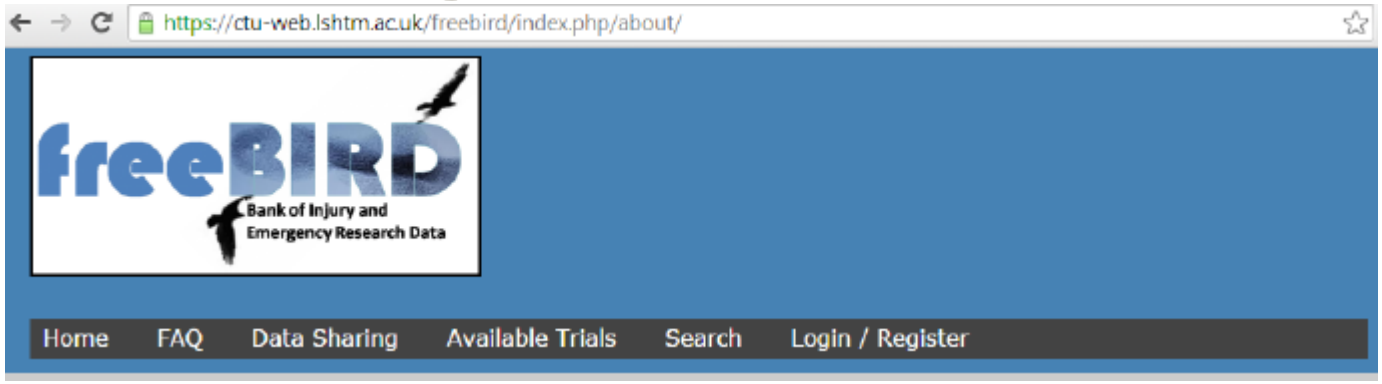
thebmj

The BMJ **requires data sharing on request for all trials**

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

All Trials Registered | All Results Reported

How to get IPD



Sandercock *et al. Trials* 2011, **12**:101
http://www.trialsjournal.com/content/12/1/101

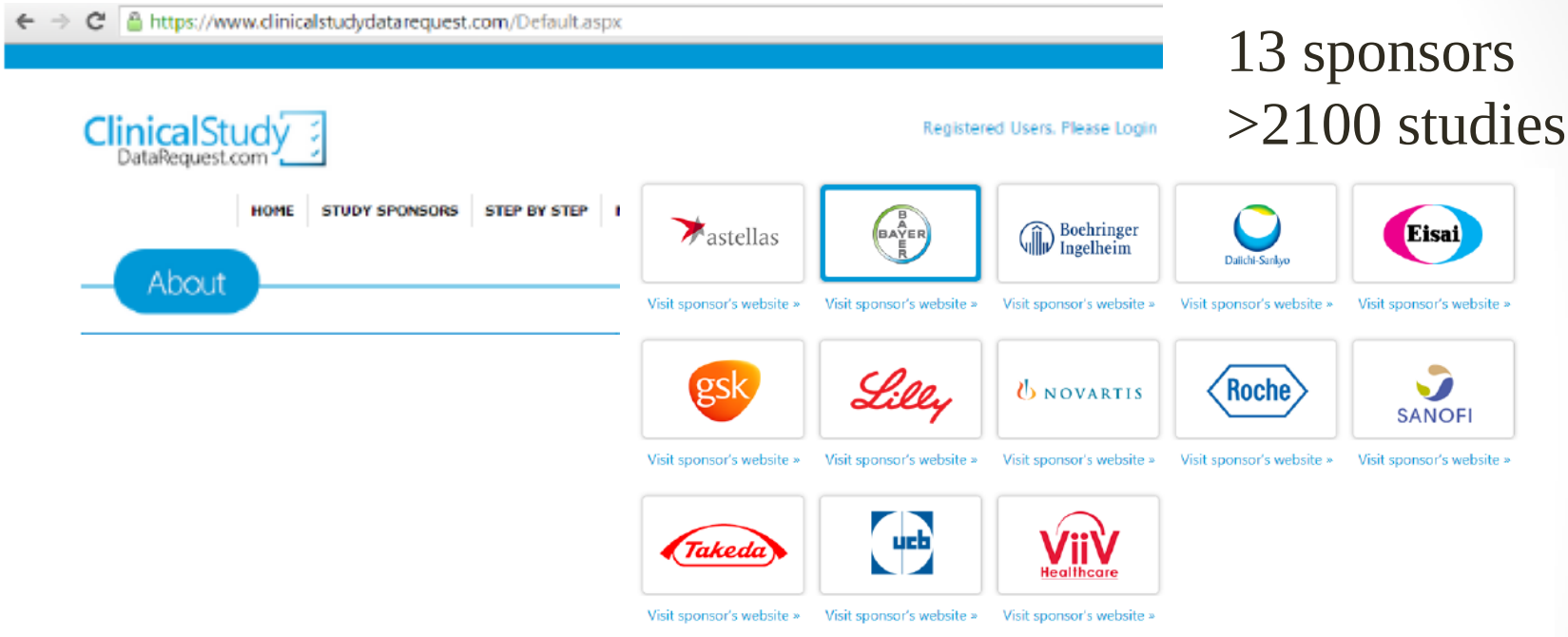
TRIALS

RESEARCH Open Access

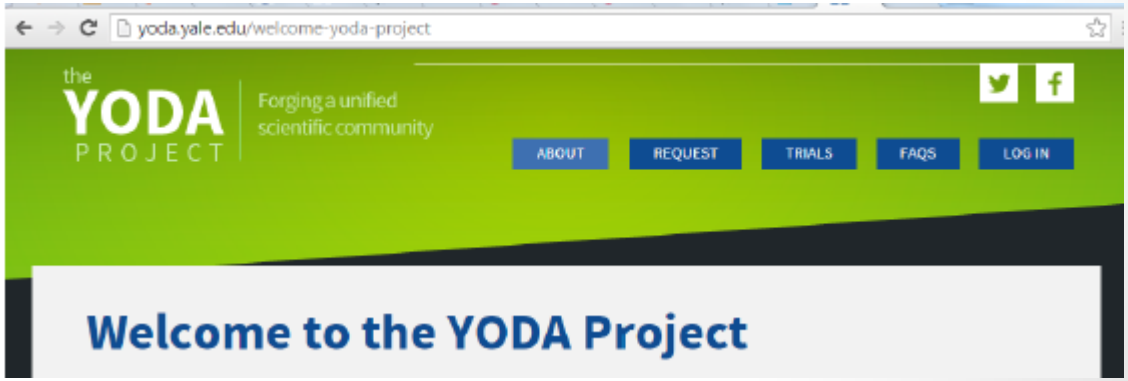
The International Stroke Trial database

Peter AG Sandercock[1*], Maciej Niewada[2,3], Anna Członkowska[2,3] and for the International Stroke Trial Collaborative Group

How to get IPD



13 sponsors
>2100 studies



2 sponsors
> 100 studies

What to do when you get the IPD

1. Understand the data (need to check the trial protocol and decipher the variable codes)
2. Replicate published results (to help with 1 and identify queries)
3. Check the data (e.g. check chronological randomisation sequence, are there any missing patients?)
4. Raise queries if possible
5. 'Clean' data
6. Recode to a consistent format across trials (depends on analysis approach)
7. Define outcomes consistently across trials
8. Analyse data - good practice to have a pre-specified statistical analysis plan
9. May need to share results with data provider

Reporting IPD meta-analysis

Special Communication

Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data
The PRISMA-IPD Statement

Lesley A. Stewart, PhD; Mike Clarke, DPhil; Maroeska Rovers, PhD; Richard D. Riley, PhD; Mark Simmonds, PhD; Gavin Stewart, PhD; Jayne F. Tierney, PhD; for the PRISMA-IPD Development Group

JAMA. 2015;313(16):1-1665. doi:10.1001/jama.2015.3656

Meta-analysis of IPD

Meta-Analysis of IPD – two stage

Stage 1: Fit model to IPD in each trial e.g for time to event data:

$$\lambda_{k(i)} = \lambda_{0(i)}(t)\exp(\beta_{(i)}x_{k(i)})$$

Treatment effect (logHR) in trial (*i*)

where $x_{k(i)}$= 1 for treatment and 0 for control for patient *k* in trial *i*

Stage 2: combine treatment effects ($\hat{\beta}_{(i)}$) and variance using standard meta-analysis method

$$\hat{\beta} = \frac{\sum_i w_i \hat{\beta}_{(i)}}{\sum_i w_i}$$

- *either fixed effect or random effects*

Meta-Analysis of IPD - one stage

combine all patient data from all studies in one single model ***taking into account the clustering of patients within study e.g. for time to event data***

Treatment effect (logHR) – assumed common 'fixed'

Fixed effect

$\lambda_{ik} = \lambda_{0i}(t)\exp(\beta x_{ik})$ where x_{ik}= 1 for treatment and 0 for control for patient *k in trial i*

Average treatment effect for a population of possible effects

Random effects

$\lambda_{ik} = \lambda_{0i}(t)\exp(\beta_i x_{ik})$

$\beta_i = \beta + b_i$ and $b_i \sim N(0, \tau^2)$

Degree of heterogeneity

Meta-Analysis of IPD - one stage

Outcome data type	Basic Model (assuming random effects)
Continuous	$y_{ik} = \propto_i + \beta_i x_{ik} + e_{ik}$ $e_{ik} \sim N(0, \sigma_i^2)$ $\beta_i \sim N(\beta, \tau^2)$
Binary	$y_{ik} \sim Bernoulli(p_{ik})$ $logit(p_{ik}) = \propto_i + \beta_i x_{ik}$ $\beta_i \sim N(\beta, \tau^2)$
Ordinal	$y_{ijk} \sim Bernoulli(q_{ijk})$ $logit(p_{ijk}) = \propto_{ij} + \beta_i x_{ik}$ $\beta_i \sim N(\beta, \tau^2)$
Count	$y_{ik} \sim Poisson(\mu_{ik})$ $ln(\mu_{ik}) = \propto_i + \beta_i x_{ik}$ $\beta_i \sim N(\beta, \tau^2)$

One-stage models

1. **Time to event outcomes** (see Tudur Smith et al Statist. Med. 2005; 24:1307–1319)
   - Fixed effect – Stratified Cox PH model
   - Random effects – SAS macro

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

3. **Binary Outcomes** (see Turner RM. et al. Stat Med 2000)
   - Generally based on logistic regression models
   - Fixed effect models - standard stats software eg SAS, R, STATA
   - Random effect models – MLwiN, Stata gllamm, winBUGS

4. **Ordinal Outcomes** (see Whitehead A. et al. Stat Med 2001)
   - Based on proportional odds models

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 vs One-stage

- Two-stage :
  - More accessible to non-statisticians
  - More in the spirit of traditional meta-analysis (can use RevMan) : Forest Plot and Heterogeneity statistics output
  - Random effects easy (not the case for one-stage time to event data)
  - Can easily incorporate both IPD and AD estimates
- But,
  - Less flexibility and more long winded
  - Lower power for detecting nonlinear associations between continuous exposures and the outcome(s) of interest
  - May lead to bias in pooled effects, standard errors, between-study heterogeneity, and correlation between random effects when few studies or few participants (or events) per study are available , when statistical models cannot fully account for follow-up times or for the time between recurrent events (see Debray et al 2015).

- **Both approaches give similar (if not identical) results most of the time!** Discrepancies can largely be explained by different assumptions rather than the number of stages (Morris and Fisher)

Software for Two Stage Approach

- Using Revman (free)

Stage 1:

Use standard statistical analysis software to obtain $\hat{\beta}_{(i)}$ - estimates of treatment effect and variance within each trial

Stage 2:

Input data using Generic Inverse Variance Method in Revman

Software for Two Stage Approach

The Stata Journal (2015)
15, Number 2, pp. 369–396

Two-stage individual participant data meta-analysis and generalized forest plots

David J. Fisher
MRC Clinical Trials Unit at University College London
London, UK
d.fisher@ucl.ac.uk

Stata command ipdmetan for two-stage IPD meta-analysis of any measure of effect

- estimates random effects and heterogeneity statistics
- can include additional covariates and interactions
- can combine IPD and AD
- produces detailed and flexible forest plots

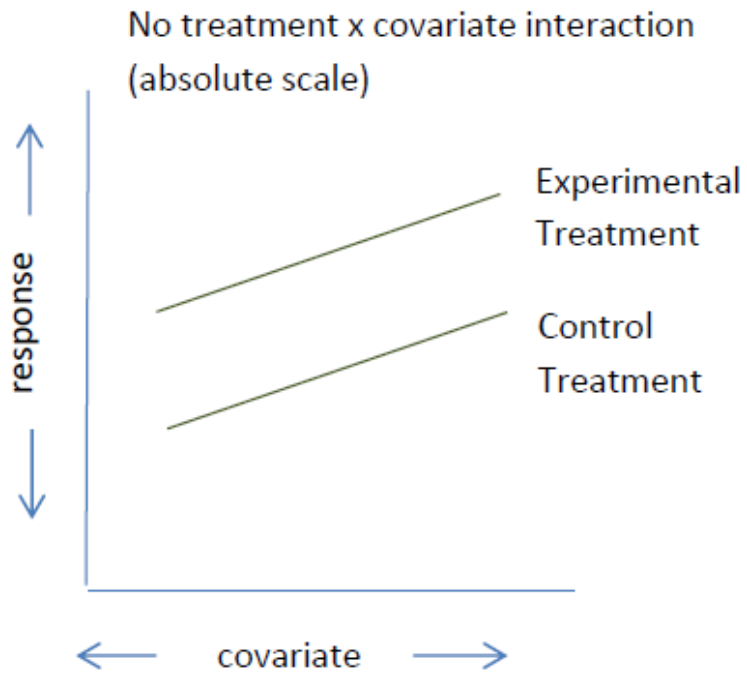
Software examples for one-stage meta-analysis

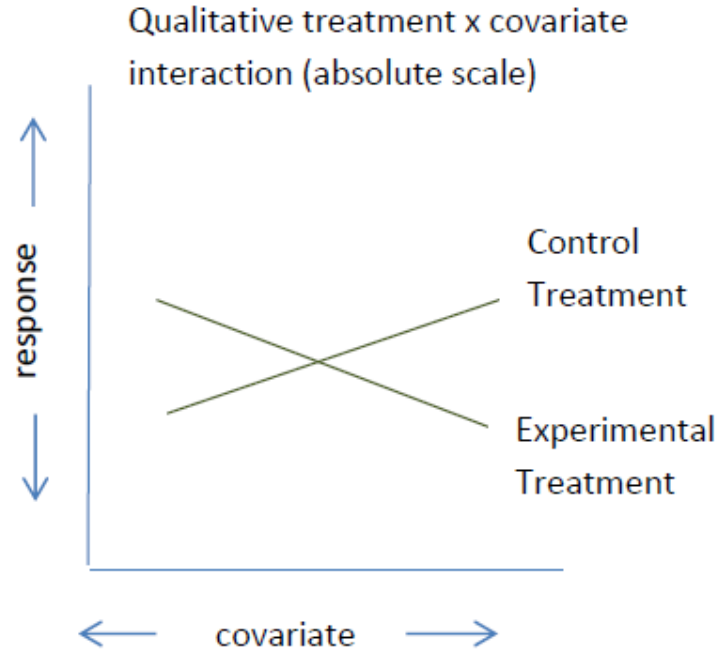
Any software that estimates multilevel mixed-effects linear models (also known as mixed-effects, multilevel, or hierarchical models)

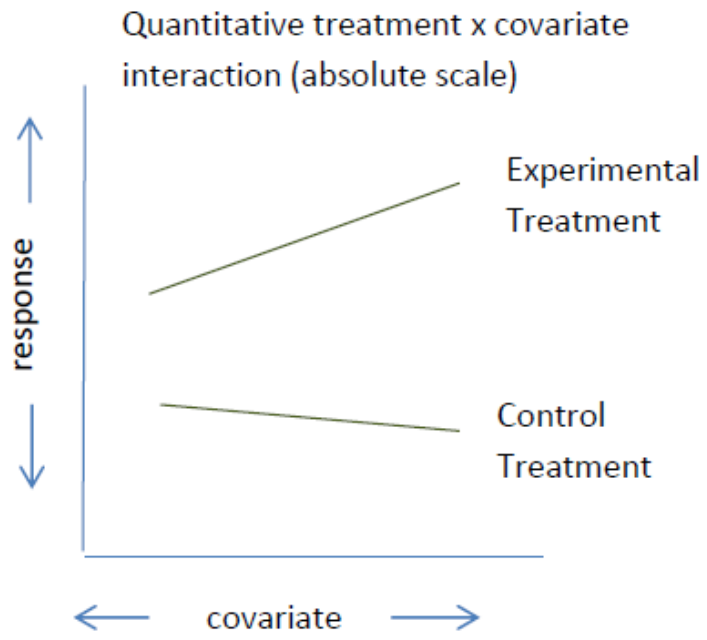
Software	command	model
R	lme4 coxme	GLMM using ML and REML (mixed linear models) Mixed effects Cox PH model
SAS	PROC MIXED	Mixed linear models using ML, REML or MOM
stata	Gllamm mixed	GLMM using ML GLMM using ML, REML and EM
MLWin	-	GLMM and survival using ML, REML and EM

For further details see Debray et al *Res. Syn. Meth.* **2015**, 6 293–309

Interactions between treatment and covariate


No treatment x covariate interaction
(absolute scale)
response
Experimental
Treatment
Control
Treatment
covariate


Qualitative treatment x covariate
interaction (absolute scale)
response
Control
Treatment
Experimental
Treatment
covariate


Quantitative treatment x covariate
interaction (absolute scale)
response
Experimental
Treatment
Control
Treatment
covariate

Investigating interactions

- Does treatment effect differ according to particular patient level characteristics?

e.g. Is carbamazepine more effective for focal seizures and valproate more effective for generalised seizures?

- Can we explain heterogeneity in treatment effects?

- To explore this we need to examine treatment – covariate interactions (also referred to as treatment effect modifier or subgroup analyses)

Investigating interactions

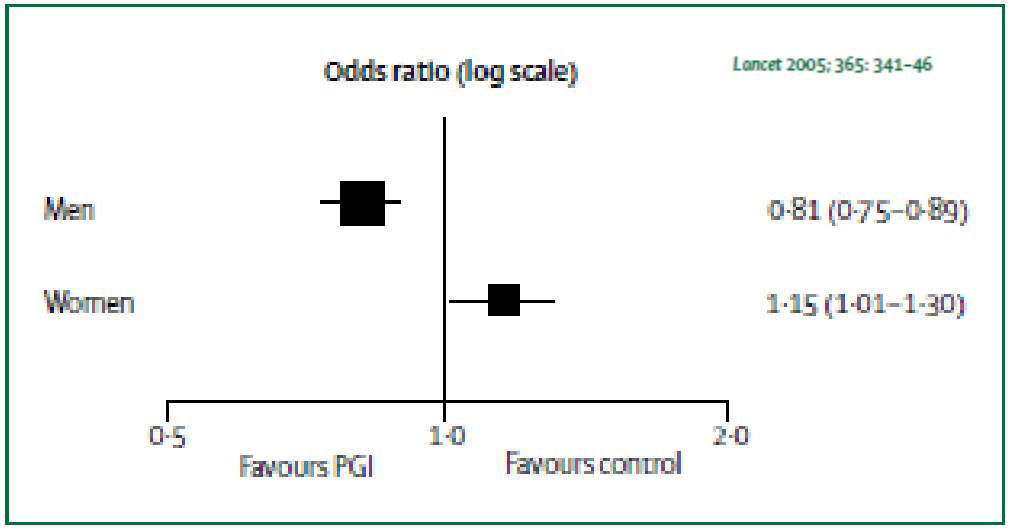


Figure 3: Odds ratios (95% CI) of death or myocardial infarction in men and women based on individual patient data

Interaction $p<0{\cdot}0001$

Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials

Lancet 2002; 359: 189–98

Why not AD ?

Detailed exploration of participant level covariates influence on treatment effect

- Meta-regression based on AD can only tell us about the across trial relationships between treatment effect and aggregated trial level covariate (eg mean age)
  - Will only identify differences if large variation in aggregated trial level covariate values
  - Ecological bias (relationship across trials doesn't necessarily reflect within trial relationship)
  - Confounding (eg an observed relationship between treatment effect and mean age may be due to higher dose of treatment given to older patients)

Why not AD ?

Detailed exploration of participant level covariates influence on treatment effect

- **Lambert et al., 2004** simulated 1000 meta-analyses, 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
  - IPD approach has a power of 90.8% to detect interactions
  - AD approach (meta-regression) has a power of 10.8% detect interactions

Investigating interactions: two stage approach

Simmonds and Higgins, 2007

Stage 1:

Estimate the treatment effect (and variance) and interaction between covariate and treatment effect, (and variance), in each trial separately

$$y_{k(i)} = \propto_{(i)} + \beta_{(i)} x_{k(i)} + \mu_{(i)} z_{k(i)} + \gamma_{(i)}\, x_{k(i)} z_{k(i)} + e_{k(i)}$$

$$e_{k(i)} \sim N\left(0, {\sigma_{(i)}}^2\right)$$

$x_{k(i)}$: treatment indicator variable (1: treated, 0: control)

$z_{k(i)}$: covariate value (eg 1: male, 0: female)

$\gamma_{(i)}$: Interaction between treatment and covariate (change in treatment effect for male compared to female)

Investigating interactions: two stage approach

Simmonds and Higgins, 2007

$$y_{k(i)} = \propto_{(i)} + \beta_{(i)} x_{k(i)} + \mu_{(i)} z_{k(i)} + \gamma_{(i)}\, x_{k(i)} z_{k(i)} + e_{k(i)}$$

$$e_{k(i)} \sim N\left(0, {\sigma_{(i)}}^2\right)$$

Stage 2:

i. Take the treatment effect estimates ($\hat{\beta}_{(i)}$) and variance for each trial and combine them in a usual fixed-effect or random-effects meta-analysis

ii. Take the interaction estimates ($\hat{\gamma}_{(i)}$) and variance for each trial (within trial estimates), and combine them in a usual fixed-effect or random-effects meta-analysis

Investigating interactions: one stage approach

Simmonds and Higgins, 2007

$$y_{ik} = \propto_i + \beta_i x_{ik} + \mu_i z_{ik} + \gamma_i x_{ik} z_{ik} + e_{ik}$$

$$e_{ik} \sim N(0, \sigma_i^2)$$

$$\beta_i \sim N(\beta, \tau^2)$$

Important:
(i) Account for clustering within trial

Assumptions about γ_i

i) Fixed (separate in each trial)
ii) Common ($\gamma_i = \gamma$)
iii) Random ($\gamma_i \sim N(\gamma, \theta^2)$)

Investigating interactions: one stage approach

Riley et al. *Statist. Med.* 2008; **27**:1870–1893

$$y_{ik} = \propto_i + \beta_i x_{ik} + \mu_i z_{ik} + \gamma_W x_{ik}(z_{ik} - m_i) + \gamma_A x_{ik} m_i + e_{ik}$$

$$e_{ik} \sim N(0, \sigma_i^2)$$

$$\beta_i \sim N(\beta, \tau^2)$$

Important:
(i) Account for clustering within trial
(ii) Separate the within and across trial interaction

Further topics

IPD unavailable…. ?

- Could studies with IPD represent a biased sample?
  - Yes if reason is related to treatment effect e.g. if IPD denied from all studies that favour control
- Can suitable AD be extracted from studies with missing IPD?
  - Undertake separate analysis of AD and compare to IPD
  - Combine if reasonable

STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:1870–1893
Published online 11 December 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3165

Meta-analysis of continuous outcomes combining individual patient data and aggregate data

Richard D. Riley[1,*,†], Paul C. Lambert[2], Jan A. Staessen[3], Jiguang Wang[4], Francois Gueyffier[5], Lutgarde Thijs[3] and Florent Boutitie[6]

NMA of IPD

Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials

Stat Med. 2013 Mar 15;32(6):914-30.

Sarah Donegan,[a*†] Paula Williamson,[a] Umberto D'Alessandro,[b] Paul Garner[c] and Catrin Tudur Smith[a]

Network meta-analysis of individual and aggregate level data‡

Jeroen P. Jansen*†

Res. Syn. Meth. **2012**, 3 177–190

References 1

- Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias and unavailable data in meta-analyses using individual participant data: a database survey BMJ 2012;344:d7762
- Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21(3):371-87
- Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RH, Reitsma JB; GetReal methods review group. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015 Dec;6(4):293-309.
- Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. Stat Med. 2013 Mar 15;32(6):914-30
- Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG: Meta-analysis of continuous outcome data from individual patients. Stat Med 2001; 20: 2219-4
- Jansen JP. Network meta-analysis of individual and aggregate level data. Res Synth Methods. 2012 Jun;3(2):177-90
- Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;55(1):86-94.
- Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. J Research Synthesis Methods 2010

References 2

- Riley RD, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008; **27**: 1870-93
- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: conduct, rationale and reporting. BMJ 2010; 340: c221
- Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007 Jul 10;26(15):2982-99
- Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF; PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA 2015 Apr 28;313(16):1657-65.
- 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.
- Stewart L, Tierney J, Burdett S. Do Systematic Reviews Based on Individual Patient Data Offer a Means of Circumventing Biases Associated with Trial Publications? In: Rothstein HR, Sutton AJ, Borenstein M, editors. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester, UK.: John Wiley & Sons, Ltd., 2006.
- Tudur-Smith C, Williamson PR, Marson AG: Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 2005, **24**: 1307-1319.
- Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG: A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000, **19**: 3417-3432
- Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. Stat Med. 2001 Aug 15;20(15):2243-60

Practical

- Undertake a two-stage and one-stage meta-analysis of IPD in R (see separate worksheet)

Please do contact me for further information

cat1@liv.ac.uk