IPD Meta-Analysis for Prognosis Research

Professor Richard Riley

Professor of Biostatistics
School of Medicine, Keele University, UK.

E-mail: r.riley@keele.ac.uk

Twitter: @Richard_D_Riley

Aims of this talk

- Introduce the concepts of meta-analysis & the use of IPD
- Rationale for embarking on an IPD meta-analysis project, rather than a traditional meta-analysis of aggregate data
- Advantages & challenges
- Notable examples
- Power (if time)

Part 1:

Traditional systematic review & meta-analysis framework using aggregate data

Meta-analysis using aggregate data

- Traditional systematic reviews & meta-analyses use **aggregate data**

- Obtainable (extracted) from study publications or study authors

e.g. reviews of randomised trials evaluating a treatment effect will extract information about participant characteristics (e.g. mean age, proportion female), study design and analysis methods, outcomes (e.g. proportion who died in each group) and key results such as:

an estimate of the treatment effect

e.g. odds ratio, relative risk, hazard ratio etc

the standard error (or variance/95% CI) of this estimate

e.g. standard error of log hazard ratio

Meta-analysis using aggregate data

- Example of aggregate data from 10 randomised trials evaluating the effect of anti-hypertensive treatment

Trial ID	Number of participants		Mean age (years)		Mean SBP before treatment (mmHg)		Mean SBP at 1 year (mmHg)		Treatment effect on SBP at 1 year adjusted for baseline (treatment minus control)
	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Estimate (variance)
1	750	704	42.36	42.17	153.05	153.88	139.75	132.54	-6.53 (0.75)
2	199	138	69.57	69.71	191.55	188.30	179.89	164.67	-13.81 (4.95)
(rows for trials 3 to 9 omitted for brevity)									
10	2297	2398	70.21	70.26	173.94	173.75	165.24	154.87	-10.26 (0.20)

Meta-analysis using aggregate data

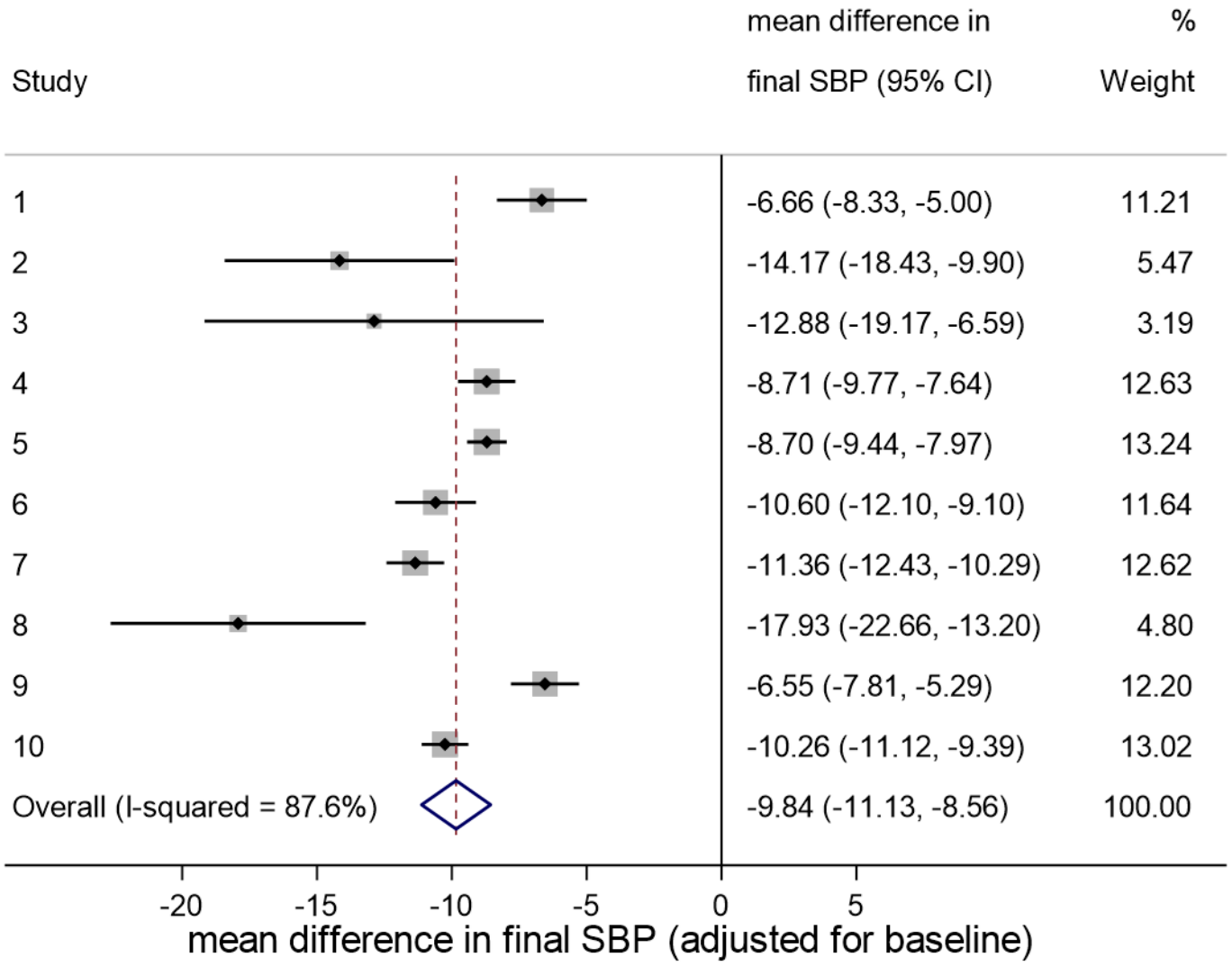
Advantages:

- Aggregate data 'simply' needs extracting (**quick** in theory, if studies are clearly and completely reported)
- Relatively **cheap** (compared to new trial; no new data collection)
- Meta-analysis methods well established:

  - such as inverse-variance common-effect or random-effects models (more later)

- Software suitable: e.g. RevMan, metafor in R, (ad)metan in Stata
- Leads to nice graphical displays such as forest plots

Example: meta-analysis of 10 hypertension trials



Meta-analysis using aggregate data

Disadvantages:

- Reliant on reporting of published articles
- Often face poor reporting (e.g. p-values rather than estimates)
- **Not in control of the statistical analysis method used**
  - Inconsistency in choice of effect (hazard ratio, odds ratio, etc.)
  - Inconsistent or no adjustment for prognostic factors
  - Complexities ignored (e.g. clustering, non-proportional hazards, non-linear relationships) etc
- Vulnerable to publication bias: studies with significant results more likely to be published (or reported well) than non-significant studies
- Vulnerable to **outcome reporting bias** – studies report only those outcomes that were significant or most interesting

Meta-analysis using aggregate data

Disadvantages:

- Going beyond original analyses is very hard (often impossible), e.g. couldn't examine proportional hazards, develop a prediction model, etc
- **Aggregate data collapses participant-level information**
  - Observe study-level summaries, such as mean age, proportion male, overall treatment effect

  Loses power to explain participant-level variation

  - Cannot adjust for prognostic factors
  - Cannot identify subgroup results, treatment-covariate interactions (effect modifiers), etc.

    i.e. can't examine whether some patients do better than others

Part 2:

IPD meta-analysis: rationale & advantages

Call for IPD meta-analysis

IPD: Individual Patient Data , Individual Participant Data

(the latter is now being adopted, as more inclusive)

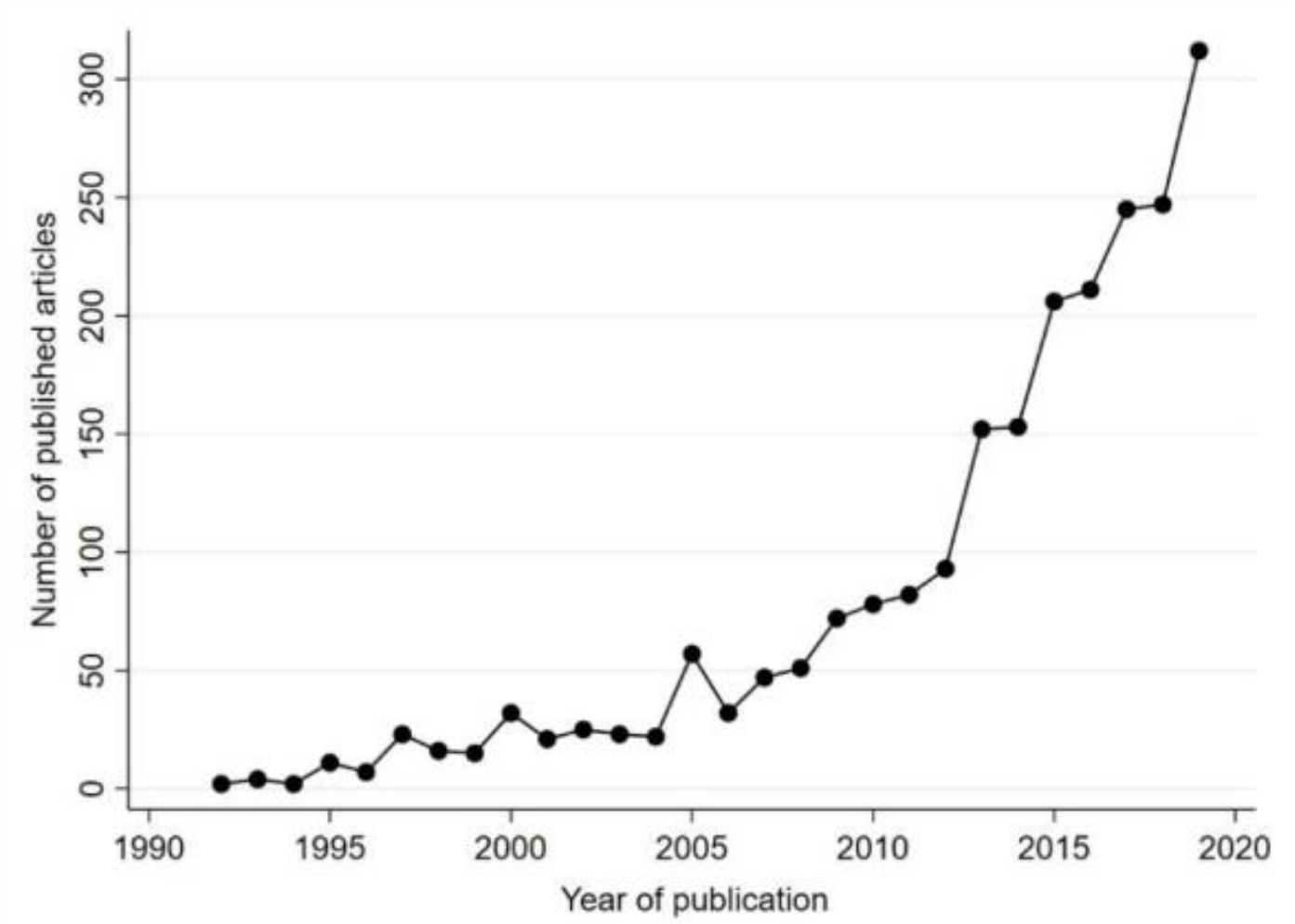
- The original, raw individual-level data from the primary studies identified by the review
- The original source material, from which aggregate data are derived

IPD meta-analysis:

The synthesis (in a statistical model) of the IPD from multiple studies for the purpose of summarising the evidence

- Increasingly relevant with the advent of 'stratified medicine' – the tailoring of treatment decisions for individual patients

Number of IPD meta-analysis articles over time (Riley, Tierney, Stewart. 2021)



Example: IPD from multiple trials, merged into a single dataset ready for meta-analysis

Trial ID	Participant ID	Treatment group, 1 = treatment 0 = control	Age (years)	SBP before treatment (mmHg)	SBP at 1 year (mmHg)
1	1	1	46	137	111
1	2	1	35	143	133
(other rows for trial 1 omitted for brevity)					
1	1454	0	62	209	219
2	1	0	55	170	155
2	2	1	38	144	139
(other rows for trial 2 omitted for brevity)					
2	337	1	44	153	129
(rows for trials 3 to 9 omitted for brevity)					
10	1	0	71	149	128
10	2	1	59	168	169
(other rows for trial 10 omitted for brevity)					
10	4695	0	63	174	128

Example: IPD from multiple cancer prognosis studies merged ready for meta-analysis

		Marker levels				Adjustment factors			Survival & disease status		
study	Patient	TH	LDH	MYCN	…	Age	Stage	…	Time of recurrence	Final survival time	Final disease status
1	1	Pos	200	5		3 yrs	1		-	150 days	ALIVE
1	2	Neg	350	3		2 yrs	4		330 days	390 days	DEAD
1	3	Neg	120	1		2 yrs	3		230 days	250 days	ALIVE with disease
2	1	Neg	320	1		6 yrs	4		27 days	48 days	DEAD
…	…	…	…	…		…	…		…	…	…

Meta-analysis using IPD

Potential advantages:

- Use consistent inclusion and exclusion criteria across studies, and if appropriate reinstate individuals into the analysis who were originally excluded
- Observe and account for missing data at the individual-level
- Verify results presented in the original study publications (assuming IPD provided can be matched to that IPD used in the original analyses)
- Inform risk of bias assessments: for example, in regard to whether groups were balanced at baseline

Meta-analysis using IPD

Potential advantages:

- Use up-to-date follow-up information
  - potentially longer than that used in the original study publications
- Identify studies which contain the same or overlapping sets of participants
- Calculate and incorporate results for those missing or poorly reported outcomes and summary statistics across published studies
  - may reduce the problem of selective within-study reporting (e.g. of outcomes)
- Calculate and incorporate results for unpublished studies
  - may thus reduce the problem of publication bias

Meta-analysis using IPD

Potential advantages:

- Standardise the strategy of statistical analysis across studies
  - e.g. the analysis method, how continuous variables are analysed, etc.)
  - use more appropriate/advanced methods than primary studies where necessary
- Assess model assumptions in each study
  - e.g. proportional hazards in Cox regression model
- Produce estimates adjusted for prognostic factors
  - may increase power, reduce heterogeneity & allows conditional treatment effects
- Adjust for a more consistent set of prognostic factors across studies

Meta-analysis using IPD

Potential advantages:

- Obtain meta-analysis results for specific subgroups of participants, and assess differential (treatment) effects across individuals
  - this facilitates individualised or stratified medicine

- Examine and compare accuracy of tests at multiple thresholds

- Generate and validate prognostic/prediction models (risk scores), and examine multiple individual-level factors in combination
  - e.g. multiple biomarkers and genetic factors, and their interaction

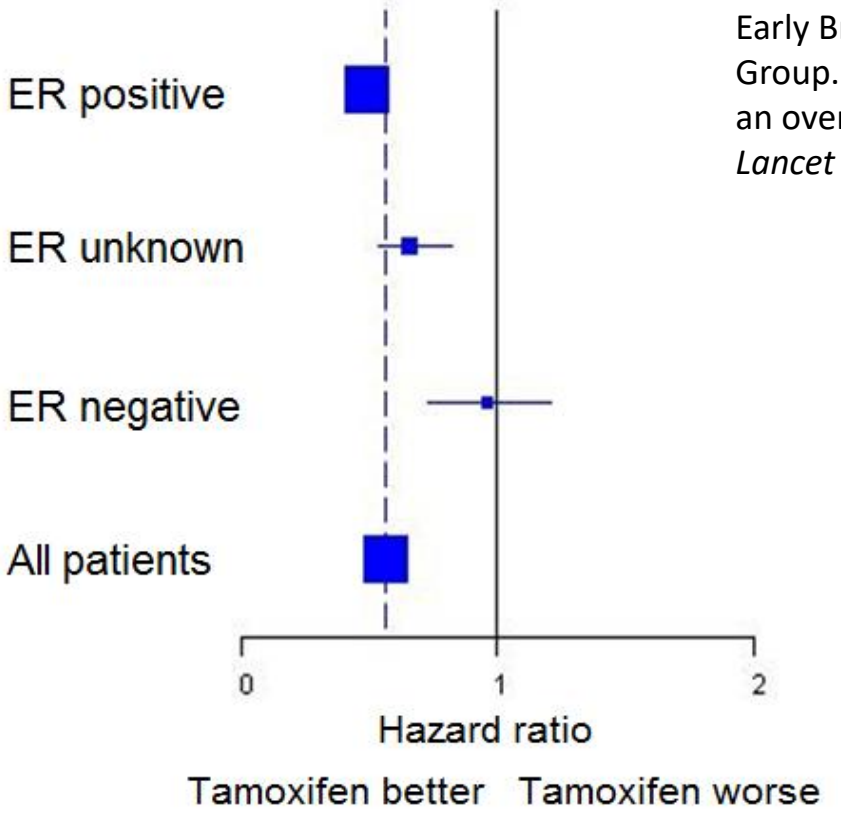
- Account for the correlation between multiple endpoints
  - a meta-analysis of longitudinal data where each participant provides results at multiple time-points

Part 3:

IPD meta-analysis projects for prognosis research: notable examples

Identification of a subgroup effect

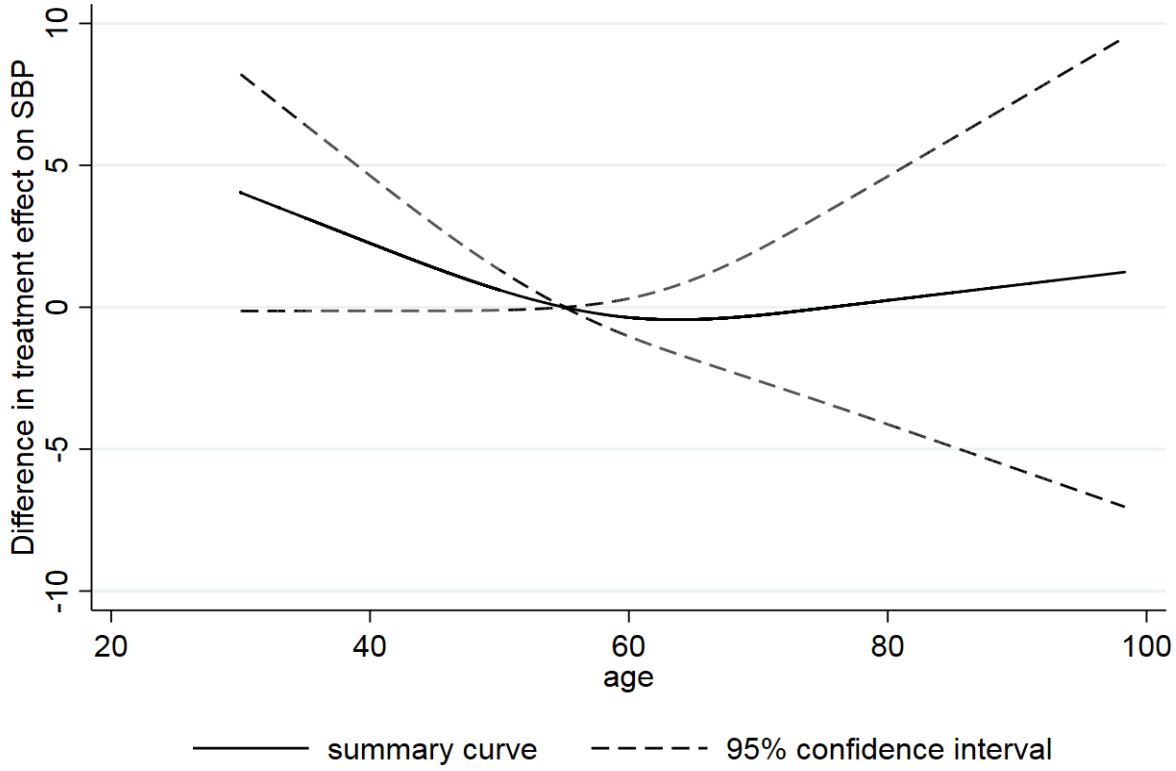
- The Early Breast Cancer Trialists' Collaborative Group obtained IPD from 55 trials, including 37000 women with early stage breast cancer.
- Examined whether the benefit of adjuvant tamoxifen varied according to oestrogen receptor (ER) status.
- Strong evidence of a larger treatment effect for the ER positive group.



Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998 351:1451-67

Non-linear relationships

- Wang et al., and then Riley et al., use IPD from 10 randomized trials to examine whether the effect of anti-hypertensive treatment differs according to age.
- IPD allows non-linear interaction to be examined - compared to those aged 55, younger patients benefit less than older benefits



Wang JG, Staessen JA, Franklin SS, et al. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. Hypertension 2005;45(5):907-13

Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. Stat Med 2020;39(15):2115-37

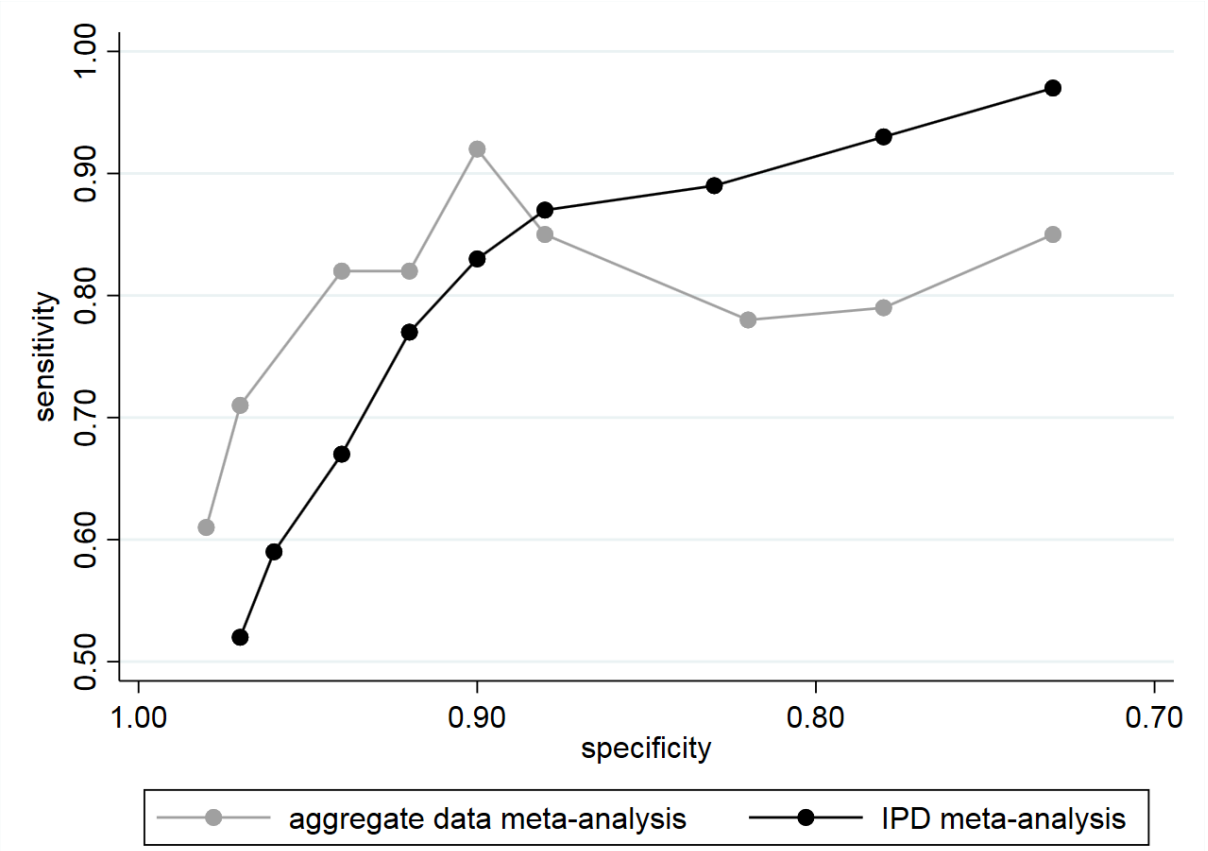
Test accuracy at multiple thresholds

- For continuous tests, different studies (selectively) report results at different thresholds
- This leads to different studies per threshold
- IPD allows any threshold to be examined in all studies and a proper ROC curve to be constructed

Figure based on:

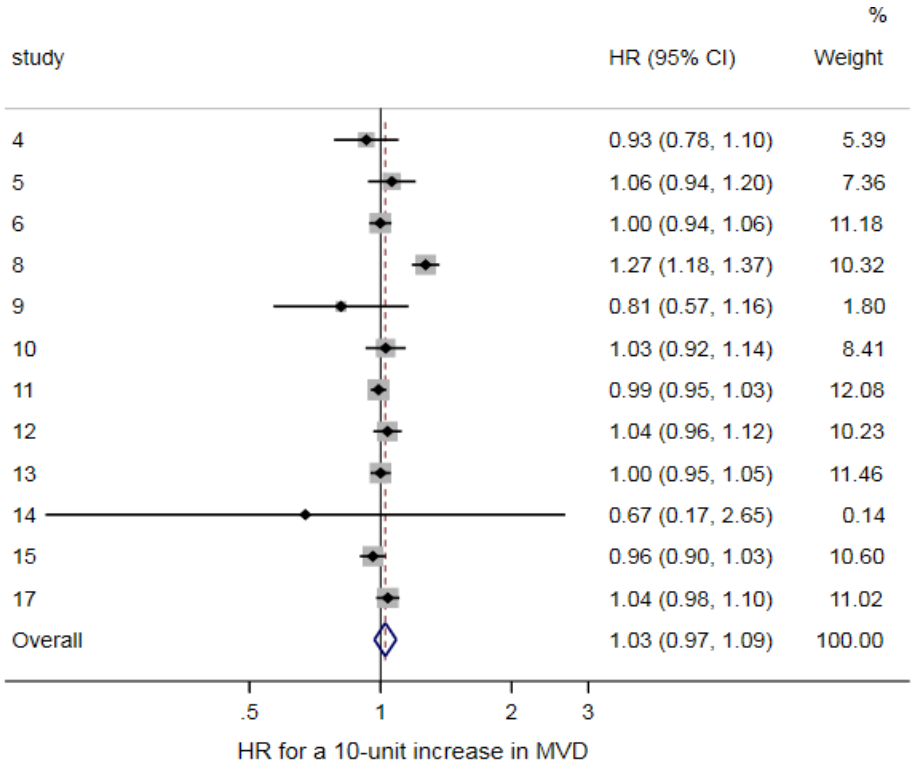
Levis B,, et al. Selective Cutoff Reporting in Studies of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool. *Am J Epidemiol* 2017;185(10):954-64

The points shown correspond to PHQ-9 threshold values of 7 to 15, from right to left.



Added Prognostic Value

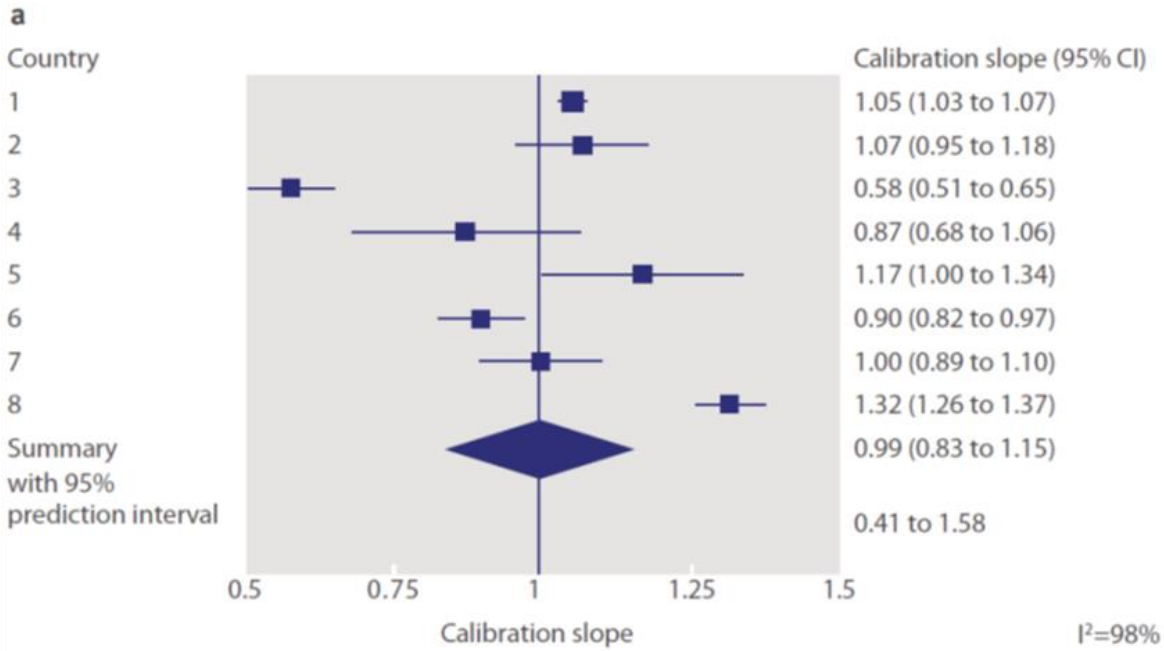
- IPD meta-analysis of IPD from 17 published and unpublished studies, involving a total of 3200 participants in non-small-cell lung carcinoma
- Is microvessel density (MVD) a prognostic factor for death?
- IPD enabled results by measurement method (here, all vessels method), adjustment for age and stage of disease, & analysis of continuous scale
- Results contradict an earlier meta-analysis using published aggregate data that concluded MVD was a prognostic effect



Trivella M, Pezzella F, Pastorino U, et al. Microvessel density as a prognostic factor in non-small-cell lung carcinoma: a meta-analysis of individual patient data. *Lancet Oncology* 2007;8(6):488-99

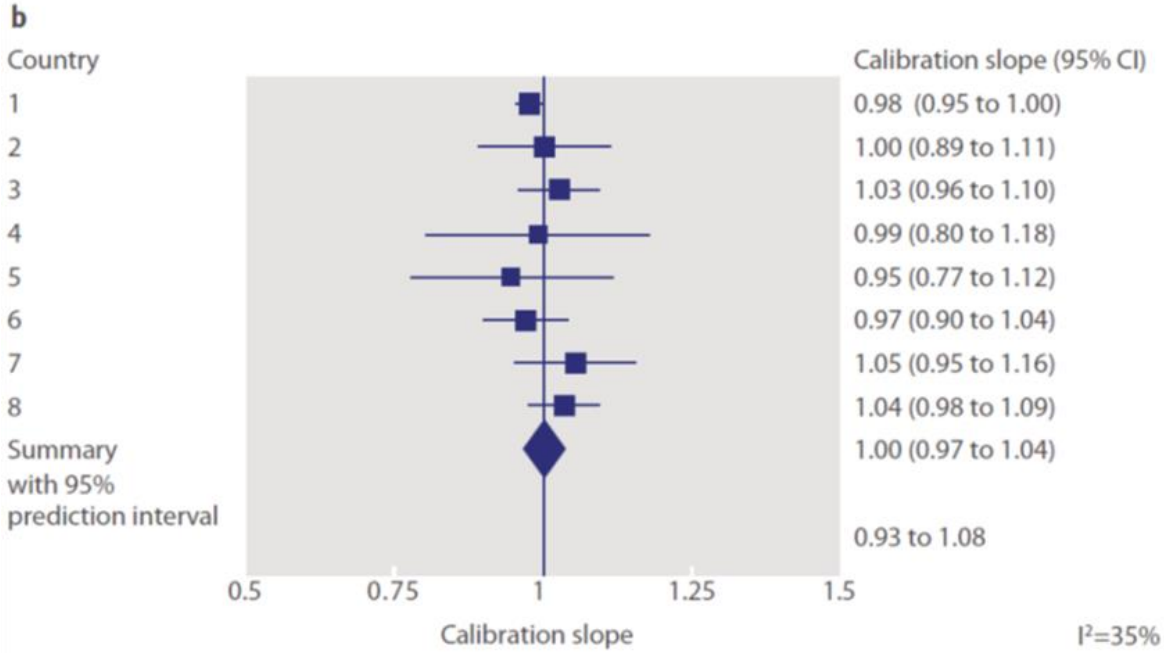
Validate & compare prediction models in multiple settings

Model 1:



Model 2:

Similar average performance but far more consistent



Riley RD, et al. *BMJ* 2016;353:i3140

Part 4:

Setting up IPD meta-analysis projects: key steps

Do I need IPD for meta-analysis?

Decision process for IPD approach:

- What is the research question?
- Has a previous review been done before to answer this question?
- What aggregate data are required to answer the question?
- Are such aggregate data available in the majority of studies?
- If not, will availability of IPD allow them to be calculated?
- How much IPD can I realistically obtain? Is it of sufficient power?
- How long will it take to obtain it?
- Do I have the resources for obtaining, collating, checking and managing large sets of IPD?
- Do I have statistical expertise and software to analyse the IPD?

Aided by: collaborating groups, different disciplines working together, leaders in the field being involved – & of course funding

Planning an IPD meta-analysis project?

- When IPD meta-analysis projects are needed, the available IPD needs to:
  - be of sufficient quality
  - record the required participant-level characteristics
  - record outcomes of interest
  - have reasonable statistical power to address the research question(s)

- Careful planning & preparatory work is needed to ensure achievable

- IPD meta-analyses are major research projects
  - **typically take upwards of two years to complete**
  - require specific research funding
  - require broader skills than conventional systematic reviews, including greater statistical expertise and experience in managing participant-level data.

Warning: Obtaining & checking IPD can be painful!

BMJ

BMJ 2013;347:f6927 doi: 10.1136/bmj.f6927 (Published 2 December 2013)

VIEWS & REVIEWS

OPEN DATA CAMPAIGN

Why did it take 19 months to retrieve clinical trial data from a non-profit organisation?

Asbjørn Hróbjartsson *The Nordic Cochrane Centre, Copenhagen, Denmark*

The emails we received during the prolonged exchange were all friendly, and the individuals involved were helpful and understood the need for data sharing, but they were hampered by inflexible, formalistic, and slow bureaucratic procedures. Since our first inquiry we communicated with four people, sent 25 emails, filled in four data use agreement forms, and waited one year and seven months.

Is your IPD project worth the effect?

• e.g. in a two-stage IPD meta-analysis, the variance of the pooled interaction (λ) is

$$\text{var}(\hat{\lambda}) = \left(\sum_{i=1}^{k} \frac{n_i \sigma_{zi}^2}{4\sigma_i^2} \right)^{-1}$$

- Studies with a larger sample size (n_i), larger variability (σ_{zi}^2) in the covariate and smaller residual variances (σ_i^2) will have larger power
- Can extract this information from trial publications

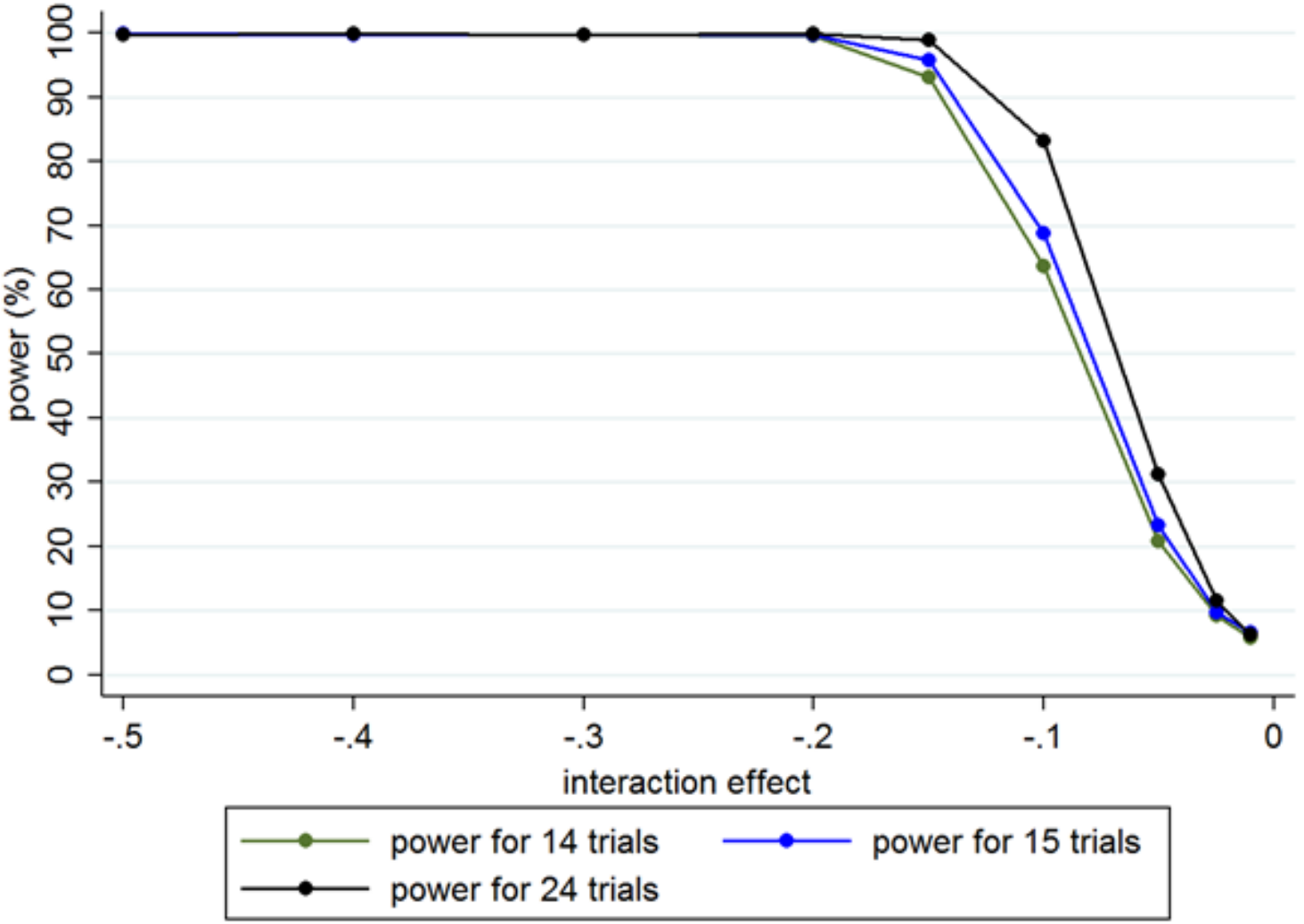
Riley RD, Debray TPA et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Stat Med* 2020;39(15):2115-37

• Assuming no heterogeneity in interaction, power is:

$$= \Phi\left(-1.96 + \lambda \sqrt{\sum_{i=1}^{k} \frac{n_i \sigma_{zi}^2}{4\sigma_i^2}} \right) + \Phi\left(-1.96 - \lambda \sqrt{\sum_{i=1}^{k} \frac{n_i \sigma_{zi}^2}{4\sigma_i^2}} \right)$$

Power example: treatment-covariate interaction

- IPD meta-analysis of 14 trials to examine if BMI is an effect modifier for interventions to reduce unnecessary weight gain in pregnancy



NEW WEBSITE www.ipdma.co.uk

INDIVIDUAL PARTICIPANT DATA (IPD) META-ANALYSIS

Home | What, why & when? | Guidance & methods | Our book | Software | Videos | Courses & resources

Promoting good practice in IPD meta-analysis projects

- entry-level information & guidance
- important steps & principles
- key articles & references
- videos of seminars & new developments
- statistical software code & information
- training courses & resources

NEW TEXTBOOK

Individual Participant Data Meta-Analysis:
A Handbook for Healthcare research

- Comprehensive introduction to IPD meta-analysis projects
- 18 chapters & over 500 pages, written and edited by researchers with substantial experience in the field
- Key concepts and practical guidance for each stage of an IPD meta-analysis project, alongside examples & learning points.
- Intended for a broad audience
- Covers trials, diagnosis, prognosis & prediction