Network Meta Analysis (NMA) Training

20/21st November 2013

James Roger, Livedata Ltd

Chrissie Fletcher, Amgen Ltd

Neil Hawkins, Oxford Outcomes Ltd / ICON

Disclaimer (Chrissie Fletcher)

- The views expressed herein represent those of the presenter and do not necessarily represent the views or practices of Amgen.

INTRODUCTION

Introducing your presenters

- James Roger
- Chrissie Fletcher
- Neil Hawkins

Introducing yourselves…

- What experience, if any, do you have relating to network meta-analysis?
- What are you hoping to get out of this course?

Course Objectives

1. Understand what is a network meta-analysis and associated terminology
2. Understand the methodology and assumptions used in a NMA
3. Understand how to plan, conduct, report and interpret a NMA
4. Gain some hands on experience of fitting an NMA
5. Understand how NMAs are used in drug development
6. Understand what reimbursement agencies think about NMAs
7. Know where to look to find more information

Structure of agenda – Day 1

- Introduction and course objectives
- What is an NMA and why is NMA important?
  - Definitions
  - Assumptions
- Steps involved in an NMA
  - Workshop
  - Case study
- NMA methodology
  - How to conduct NMA for different types of endpoints
    - continuous
  - Frequentist vs Bayesian
  - Workshop

Structure of agenda – Day 2

- NMA methodology cont.
  - How to conduct NMA for different types of endpoints
    - Binary, count, hazard ratios etc
  - Workshop
- New NMA techniques
- Applications of NMA
  - Industry perspective: drug development
  - Academic/payer perspective: evidence based medicine
  - Workshop
- NMA in HTA methodology guidelines
- NMA best practices
- Conclusions & wrap-up

WHAT IS AN NMA?

Lots of different terminology used

- Indirect comparison
- Indirect treatment comparison
- Adjusted indirect comparison
- Adjusted indirect treatment comparison
- Mixed treatment comparison
- Network meta-analysis

Definitions

Indirect (Treatment) Comparison

A comparison of treatments that have not been compared ‘head-to-head’ in a randomised controlled trial (RCT)

Adjusted Indirect (Treatment) Comparison

A comparison of the relative treatment effects using a common comparator

[These may often be referred to as an Indirect (treatment) comparison]

Definitions

Mixed treatment comparison

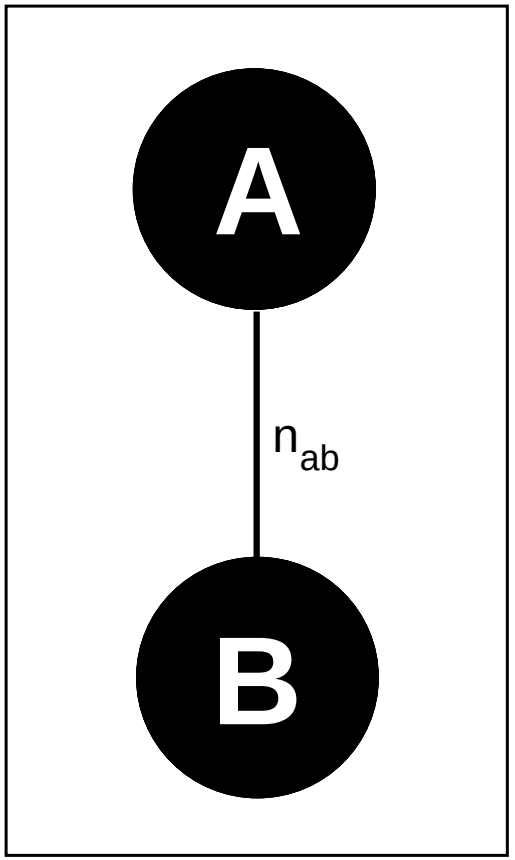
Combining treatment effects obtained from direct (head to head) RCTs with indirect estimates of treatment effects

Network meta-analysis

Allowing multiple pairwise comparisons for many treatments to be estimated simultaneously to provide relative treatment effects of multiple treatment comparisons

Direct Comparison

- One or more RCTs
- Meta-analysis of AvB trials to gain better estimate of the overall treatment difference
- Usual issues with meta-analysis apply
  - Publication Bias
  - Heterogeneity

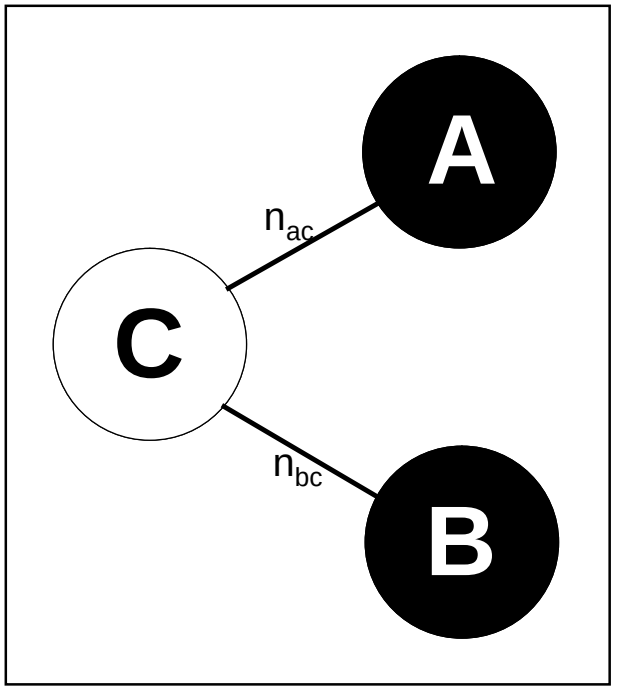


Indirect Comparison

- Indirect comparison of AvB is obtained from meta-analysis of AvC trials and meta-analysis of BvC trials

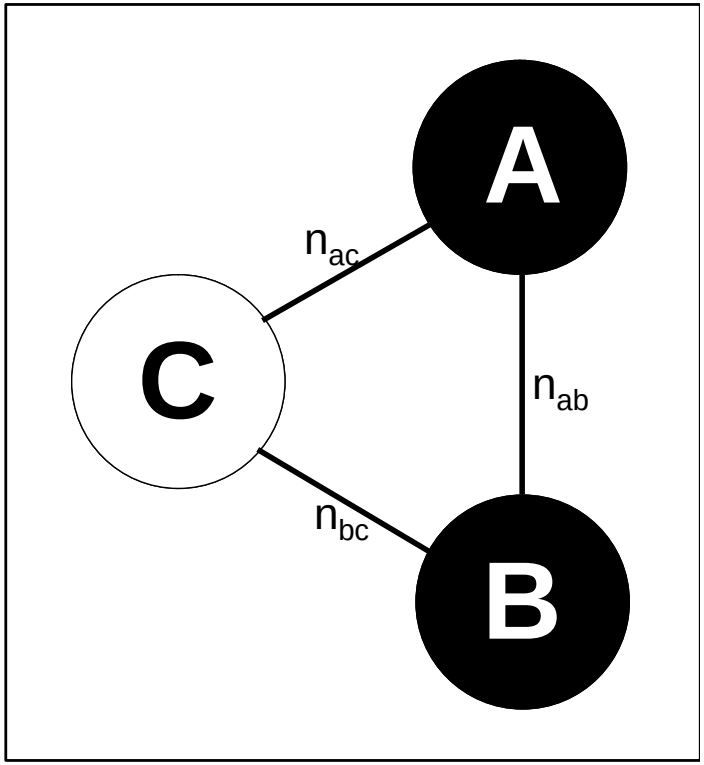
$$\text{Indirect } (\delta_{a\text{-}b}) = \delta_{ac} - \delta_{bc}$$
$$\text{or, Indirect } (\delta_{a/b}) = \delta_{ac}/\delta_{bc}$$

- Two sets of meta-analysis assumptions
- Extra assumptions are involved



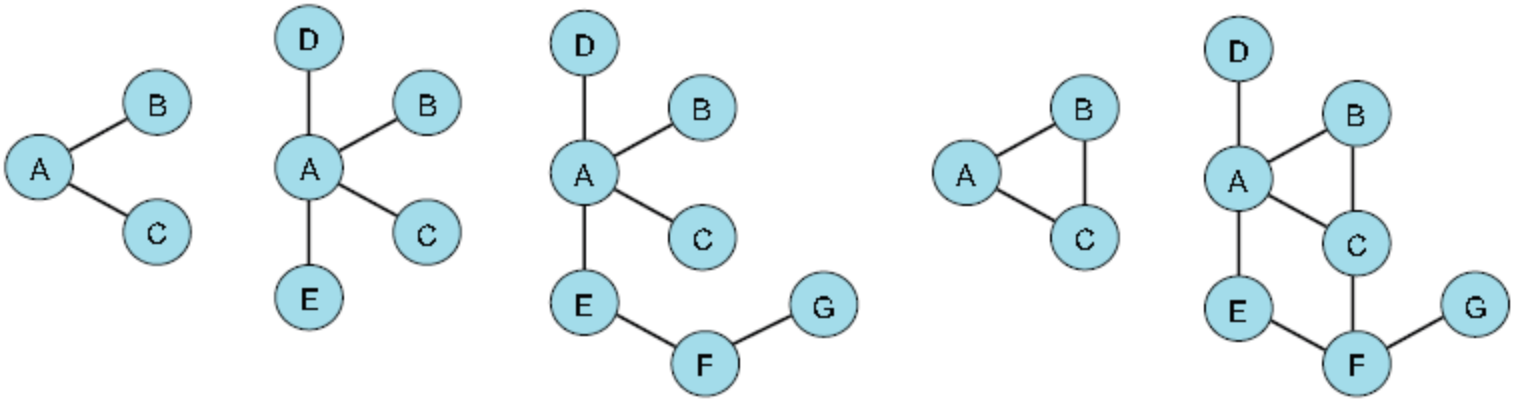
Network Meta Analysis

- Mixture of direct and indirect comparisons from meta-analyses of RCTs
- Three sets of meta-analysis assumptions
- Additional assumptions



ISPOR Indirect Treatment Comparison Good Research Practices report (part 1)*

Networks of evidence



Closed loops in network: combination of direct and indirect evidence

Definitions

Anchored Indirect Treatment Comparison
(or 'adjusted' ITC)

Mixed Treatment Comparison

Network meta-analysis
(>2 studies in network)

* http://www.ispor.org/workpaper/interpreting-indirect-treatment-comparison-and-network-meta-analysis-studies-for-decision-making.pdf

Other NMA Definitions - EUnetHTA

1.1. Definitions and general information

Direct comparison – the combination of multiple head-to-head trials to generate a pooled estimate of the relative effectiveness of the two treatments

Indirect comparison – the estimation of the relative effectiveness of two or more treatments in the absence of any head-to-head trials

Multiple treatment comparison – the estimation of the relative effectiveness of three or more treatments

Mixed treatment comparison – the simultaneous estimation of the relative effectiveness of three or more treatments using a combination of direct and indirect evidence

To compare two or more treatments, meta-analytic techniques are generally used to combine the results of multiple studies in an attempt to provide the best evidence base. A meta-analysis is the formal evaluation of the quantitative evidence from two or more studies addressing the same question. This most commonly involves the statistical combination of summary statistics from the various studies, but the term is sometimes also used to refer to the combination of raw data. Direct comparisons enable evidence synthesis based on multiple head-to-head trials. Where direct head-to-head evidence is lacking, indirect evidence can be used to supplement the relative effectiveness data from the direct comparisons available.

EUnetHTA relative effectiveness guidelines (released Mar 2013):
http://www.eunethta.eu/outputs/methodological-guideline-rea-pharmaceuticals-direct-and-indirect-comparison

ASSUMPTIONS

Assumptions

- Exchangeability
  - This is the assumption that the **direct effects ,** $\delta_{ap(i)}$, are exchangeable with all the other treatment effects in the AvP trials

$$\delta_{ap(i)} \sim N(\delta_{ap}, \sigma^2_{ap})$$

  - If this can be assumed then different contrasts of treatments that give **indirect effects** are also exchangeable

$$\delta_{ab(i)} \sim N(\delta_{ab}, \sigma^2_{ab})$$

  - Difficult to assess directly

Assumptions

- Homogeneity
  - For each pair-wise comparison, are trials clinically and statistically comparable
  - Heterogeneity can be present in the individual meta-analyses that comprise the dataset for an NMA
  - Heterogeneity could be explained by adjusting for study level baseline characteristics.
  - The influence of heterogeneity can be mitigated by using a random effects modelling approach (standard approach)
  - Can also be mitigated by moving from a ‘difference’ treatment estimate to a ‘ratio’ estimate

Assumptions

- Similarity
  - The assumption that an indirect comparison does not differ by patient subgroups
  - Thorough exploration of patient subgroup is required to show that indirect comparison estimates are not different among subgroups, or influenced by outlier studies.
  - Use of meta-regression techniques
  - Important to show that indirect comparisons are not influenced by study level patient characteristics

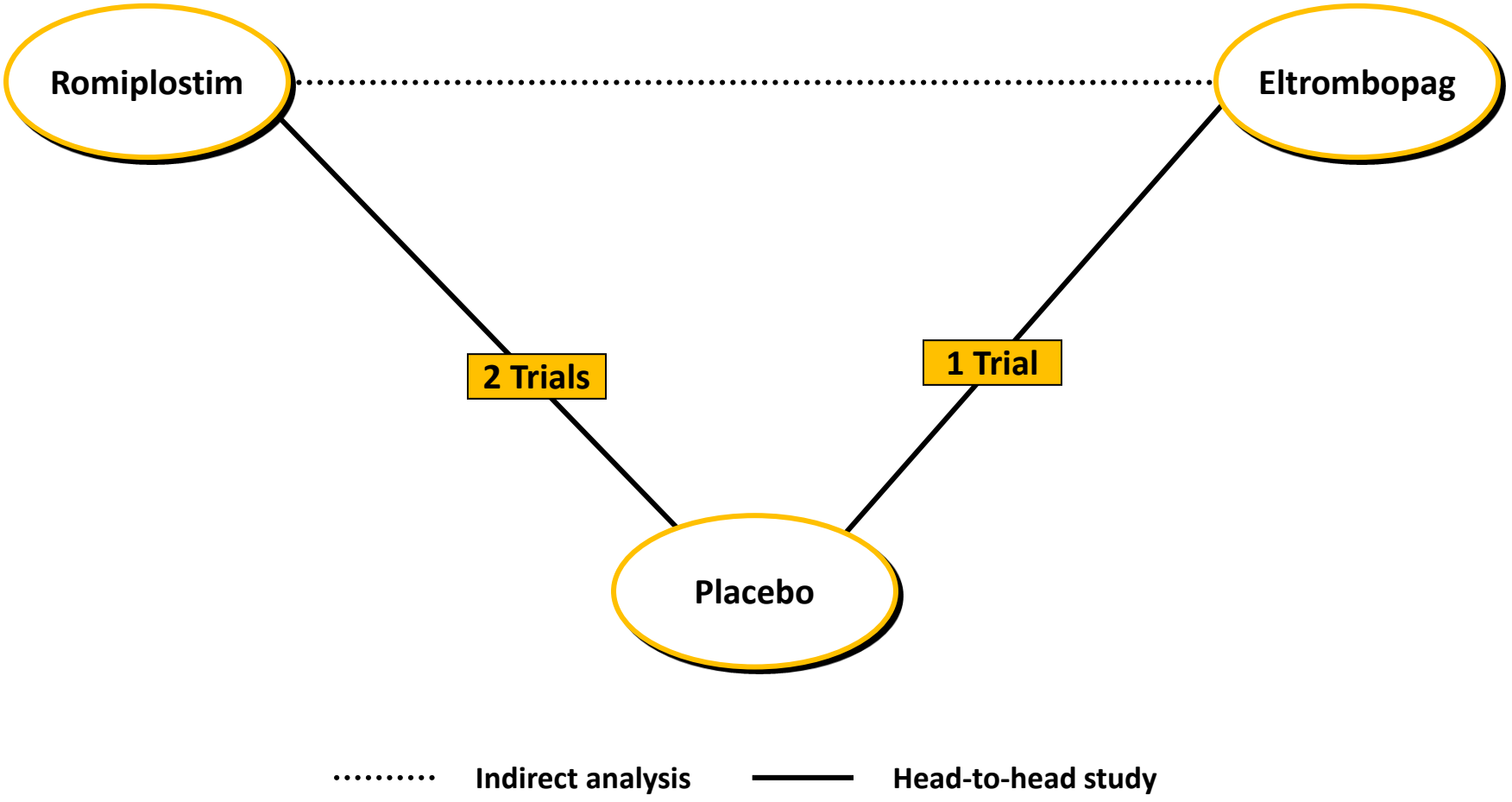
Assumptions

- Consistency
  - The assumption that the indirect evidence is consistent with any direct evidence
  - This can be explored by comparing the discrepancy between direct and indirect estimates

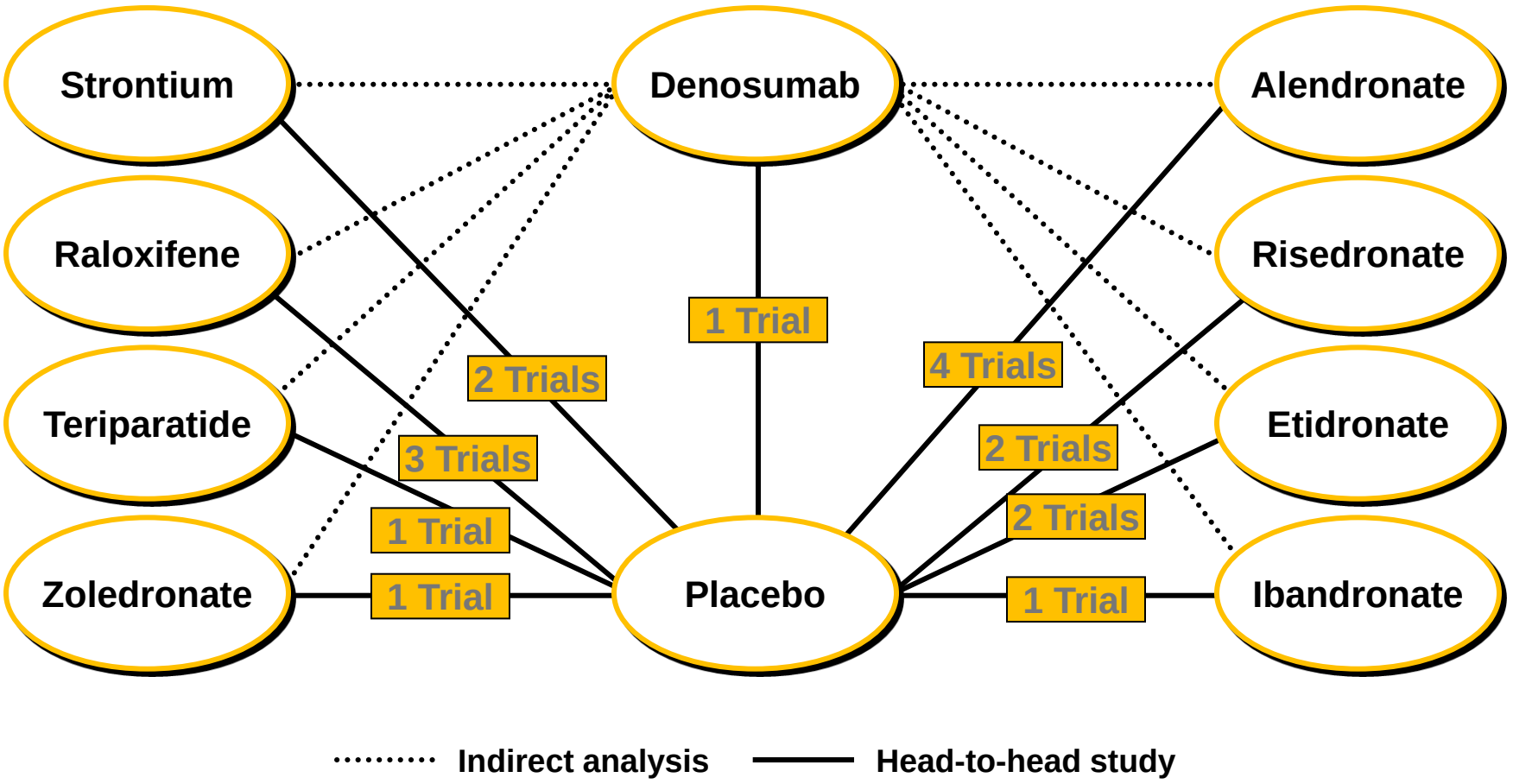
$$\text{discrepancy } \omega_{ab} = \delta^{dir}_{ab} - \delta^{ind}_{ab}$$

  - For comparison of AvB, the estimate based on only the direct evidence is compared to the estimate obtained from the NMA excluding the direct AvB evidence. This can be done for each comparison in the NMA.
  - DIC is used to detect inconsistencies
  - See also Dias et al (2010) and NICE DSU Report 4

EXAMPLES OF NMAS



Cooper KL et al. Immune Thrombocytopenia. *Int J Technol Assess Health Care.* 2012;28:249–258



Freemantle et al. Osteoporosis, International Osteoporosis Foundation and National Osteoporosis Foundation. 2012

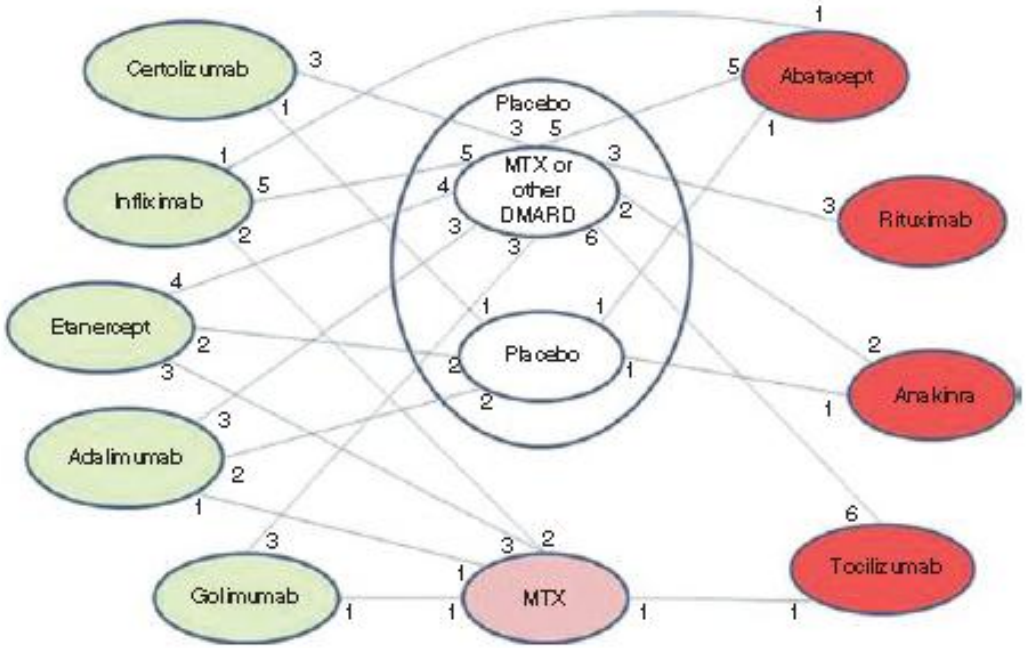


Figure 1 Network diagram of available direct comparisons. The numbers at the lines indicate the number of trials for each direct comparison. DMARD, disease-modifying antirheumatic drug; MTX, methotrexate.

Mandema. A Dose–Response Meta-Analysis for Quantifying Relative Efficacy of Biologics in Rheumatoid Arthritis

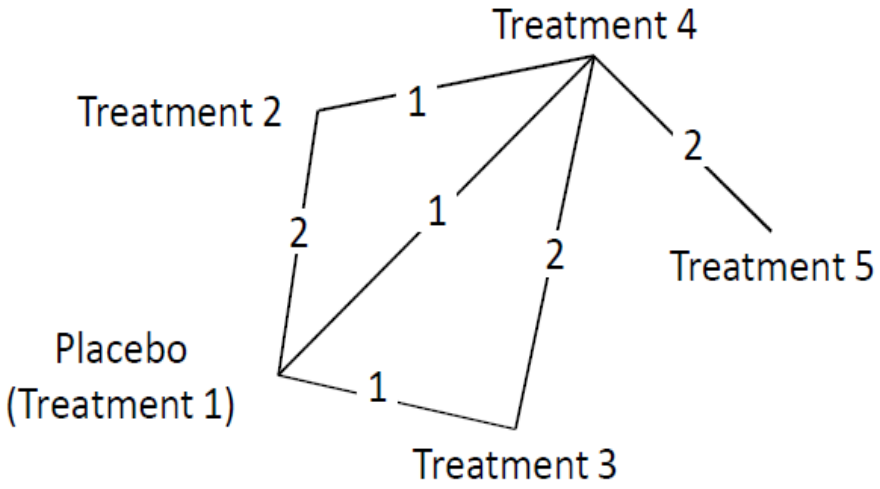
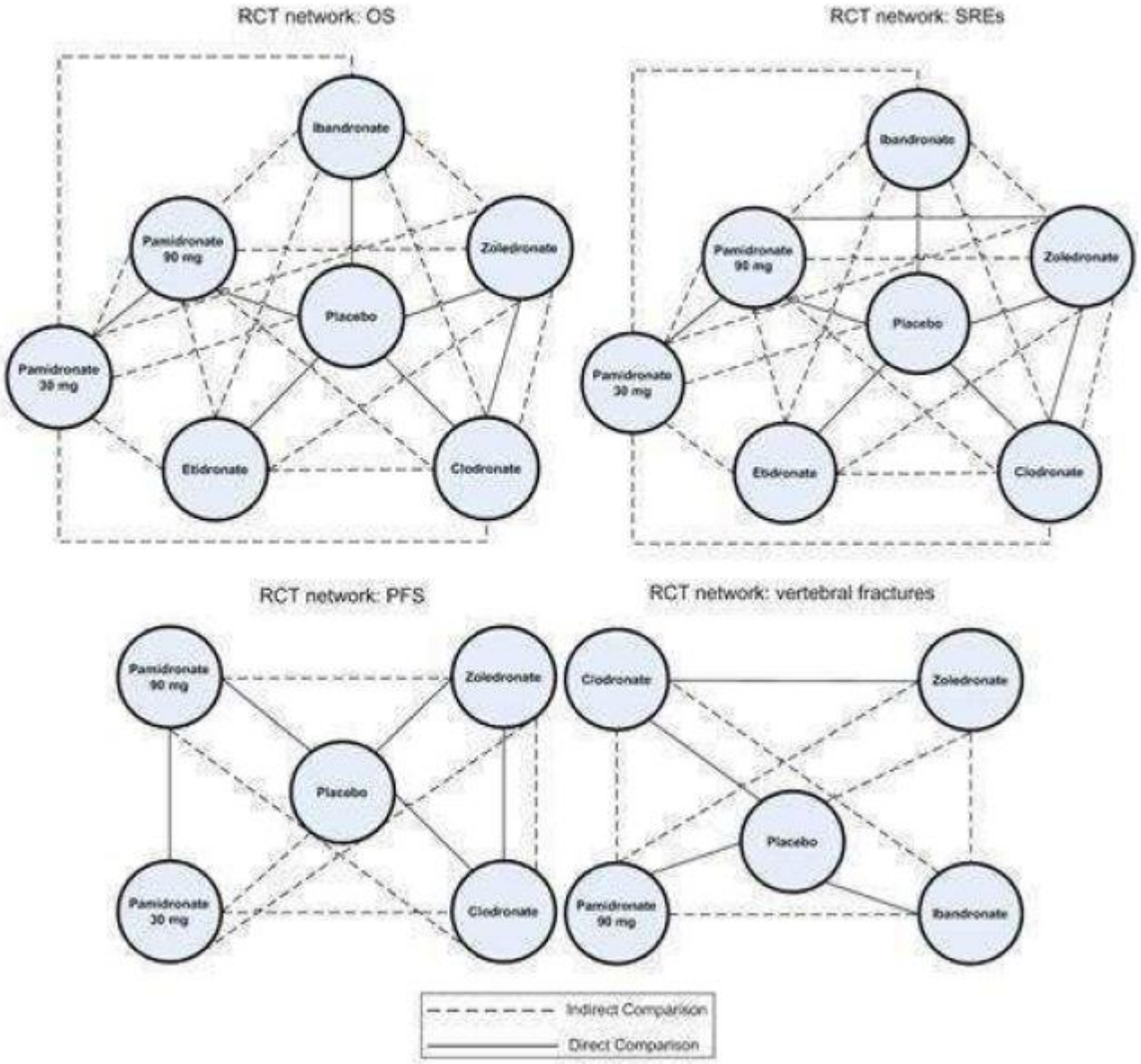


Figure 3 Parkinson network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison.

NICE Decision Support Unit Technical Series Document 1 Introduction to evidence synthesis for decision making

Figure 5. Randomized controlled trial (RCT) networks. OS: Overall survival; SREs: Skeletal-related events; PFS: Progression-free survival.



WHY IS NMA IMPORTANT?

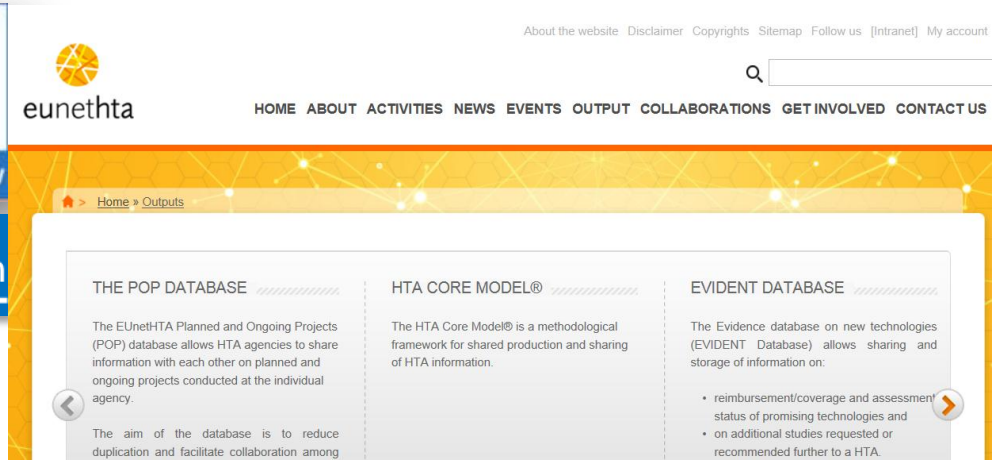
Rationale for network meta-analyses

- Increasingly many organizations are comparing new treatments against existing therapies by using network meta analysis techniques
- Questions concerning the comparative effectiveness (US) or the relative effectiveness (EU) of a new treatment receiving regulatory approval are being raised by numerous healthcare, clinical and government stakeholders.

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care www.ahrq.gov

Effective Health Care Program

EUnetHTA - European network for Health Technology Assessment



Relative Effectiveness/Efficacy HLPF & EFPIA Definition

- Referred to as Relative effectiveness*
  - Relative Effectiveness can be defined as the extent to which an intervention does more good than harm compared to one or more intervention alternatives for achieving the desired results when provided under the usual circumstances of health care practice
- Different than Relative Efficacy*
  - Relative Efficacy can be defined as the extent to which an intervention does more good than harm, under ideal circumstances, compared to one or more alternative interventions

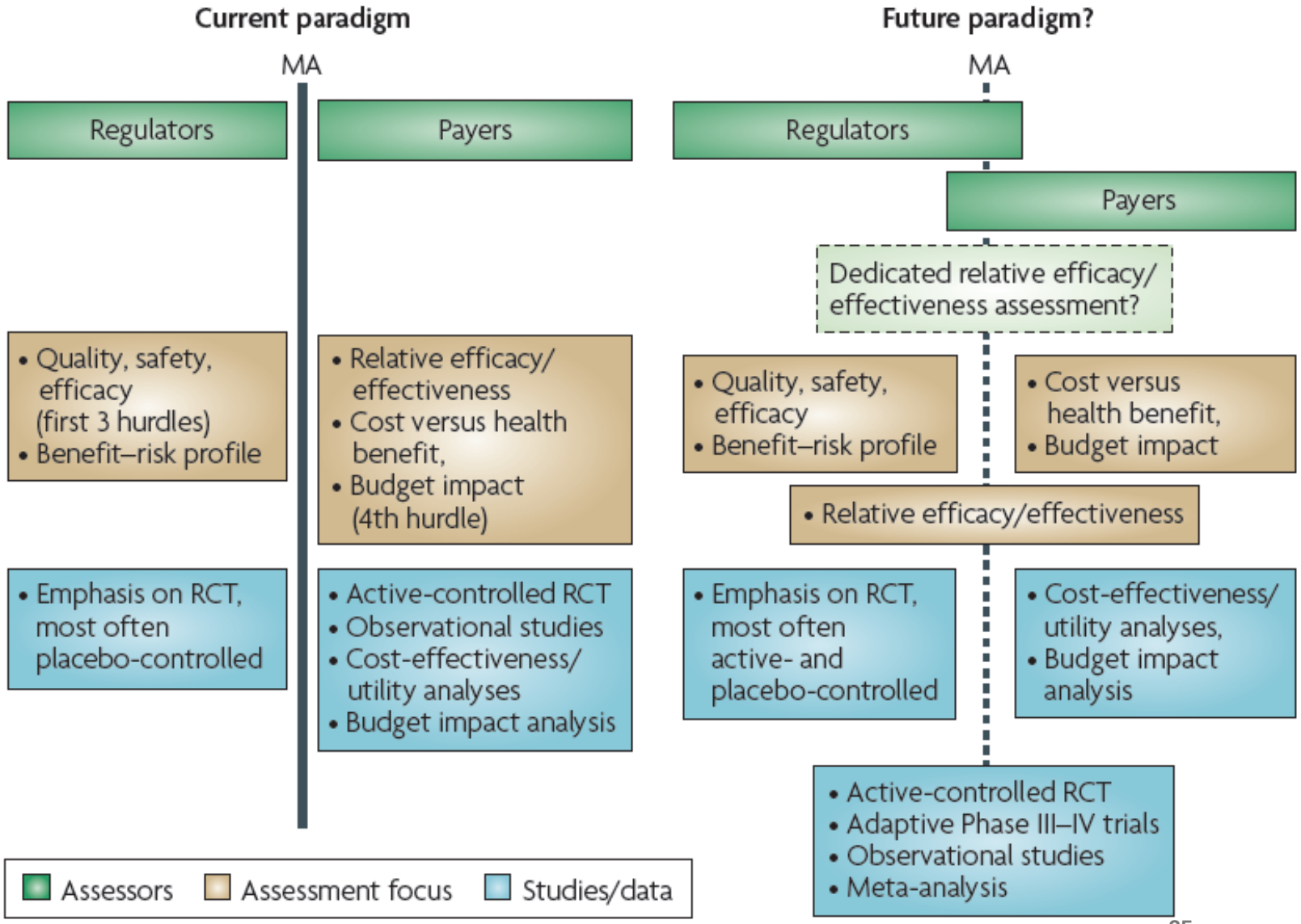
*** Adopted by High Level Pharmaceutical Forum**

Comparative Effectiveness (CE) – US Definition

- Comparative Effectiveness Research* is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions.
  - The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.
  - To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations.
  - Defined interventions compared may include medications, procedures, medical and assistive devices and technologies, behavioural change strategies, and delivery system interventions.
  - This research necessitates the development, expansion, and use of a variety of data sources and methods to assess comparative effectiveness.

*** US Dept of Health**

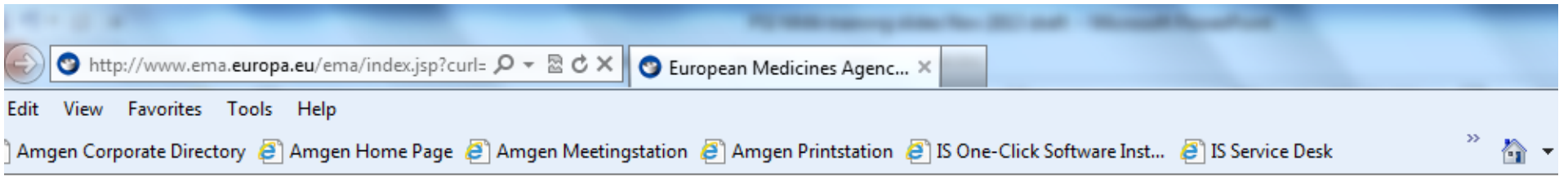
Evidence needs for regulators and payers



Eichler et al, Nature Reviews 2010

Areas of mutual interest between regulators and payers

- Exchange of information
- Parallel scientific advice
- Debate on evidence requirements
- Relative efficacy assessment
- Alignment on post-marketing research activities
- Parallel review
- Managed market entry (provisional/progressive decisions)



News and press release archive
Committee meeting reports
Calendar
Statistics
What's new
Media centre
Brochures
RSS feeds
Newsletters
Social media

European Medicines Agency and EUnetHTA review progress of their cooperation

Press release

07/06/2013

European Medicines Agency and EUnetHTA review progress of their cooperation

Focus on facilitation of development plans through advice procedures

The European Medicines Agency (EMA) and EUnetHTA, the European Network for Health Technology Assessment, met to review the progress of their cooperation in London on 14 May 2013. This was the sixth meeting since the start of their collaboration in 2010.

The focus of this meeting was on how regulators and health-technology-assessment (HTA) bodies can work together to facilitate drug development by cooperating in giving advice to pharmaceutical companies. EUnetHTA is piloting joint early dialogue with technology sponsors by a number of national HTA agencies and the EMA has a scientific

EMA website:
http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2013/06/news_detail_001807.jsp&mid=WC0b01ac058004d5c1

NMA is increasingly accepted by more HTA agencies

- Time to market has become time to reimbursement and not time to regulatory approval
- NMA becoming critical as part of evidence synthesis and demonstrating how a new treatment compares to existing therapies used in local medical practice to support a local HTA at product launch
- Increasing exchange of scientific information between European HTA networks (EUnetHTA) and payers within member states on evidence to support health care decisions
- Methodology has evolved and NMA becoming more accessible and accepted where gaps in evidence exist

Summary from HTA agency methodology guidelines

- NMAs should only be conducted when RCTs don’t exist
- Less weight is given to a NMA compared to RCTs
- Observational data should not be used in a NMA
- Most note that a NMA has relatively low power to detect important differences
- All HTA bodies comment on the underlying assumption that a NMA is only valid if the contributing RCTs are similar

Introducing the IMI* '*GetReal*' project

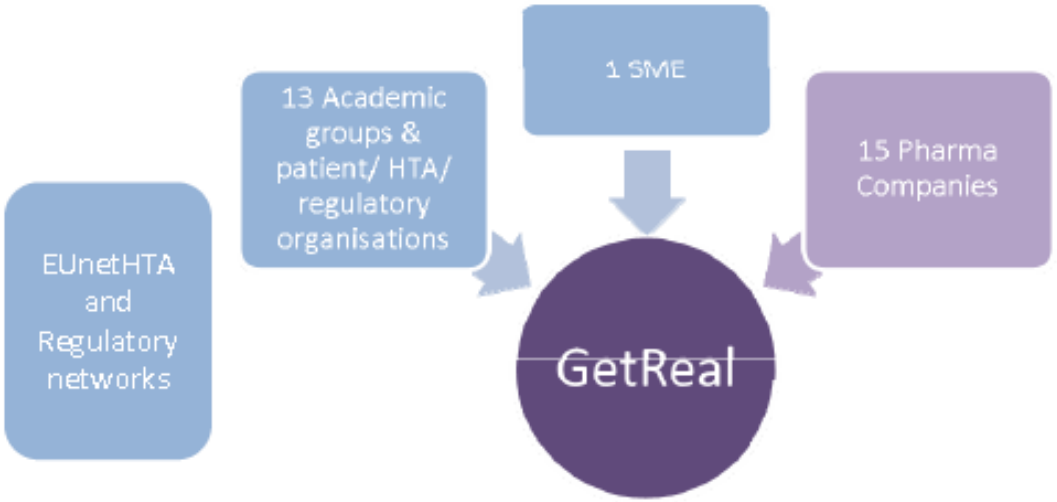
GetReal

Project Vision

For Pharmaceutical R&D and healthcare system decision makers to jointly understand how real world data and analytical techniques can best be used to improve the value of information available at **marketing authorisation:** contributing to better informed and more consistent assessments underpinning patient access to new medicines.

Lasting impact

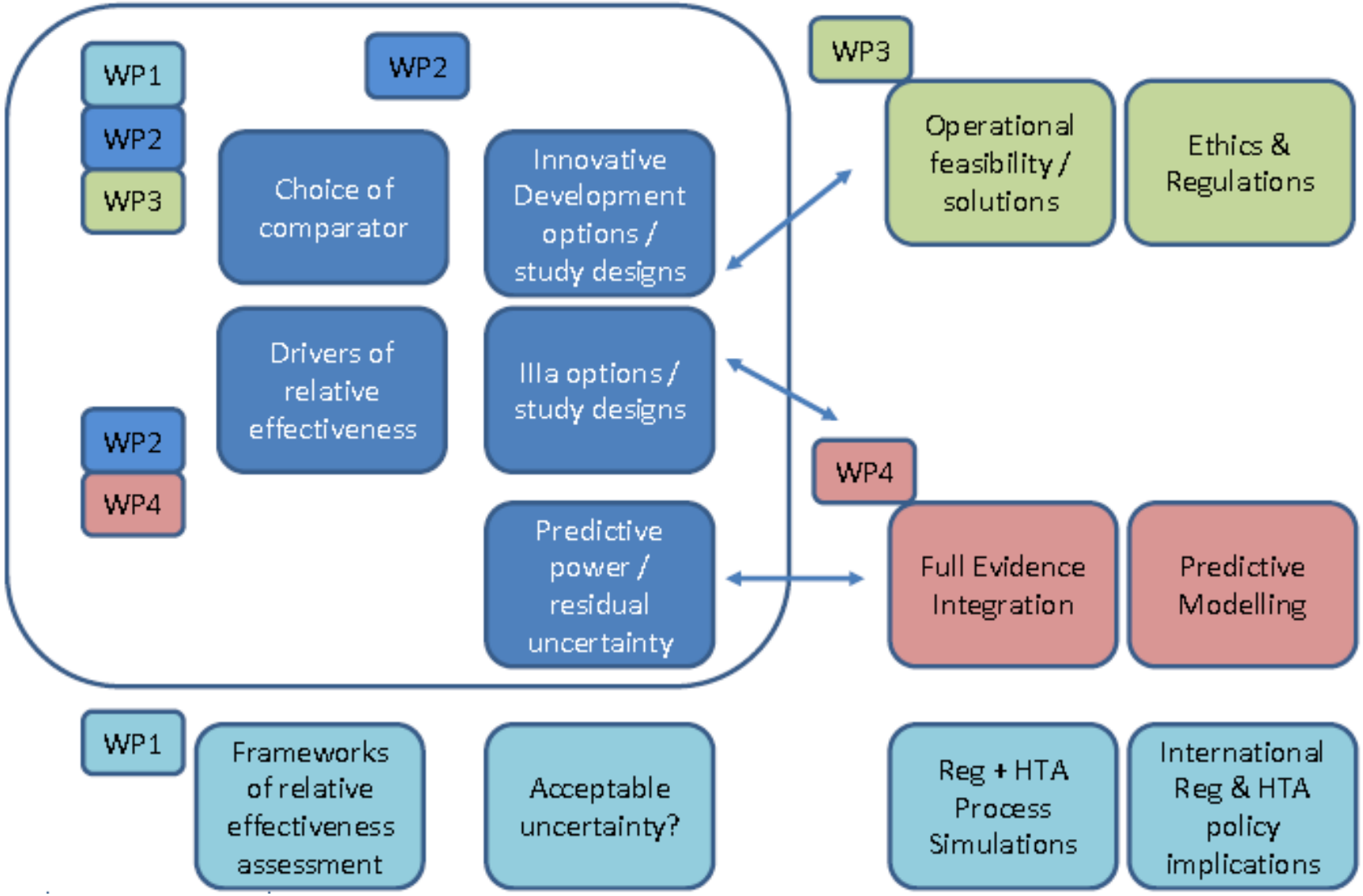
To provide a methodological and analytical framework that informs policy and process evolution beyond the life of the project and at an international level; and to provide tools, techniques and training that ensure that the potential of real world data can be exploited in drug development.



IMI GetReal: Project deliverables and benefits

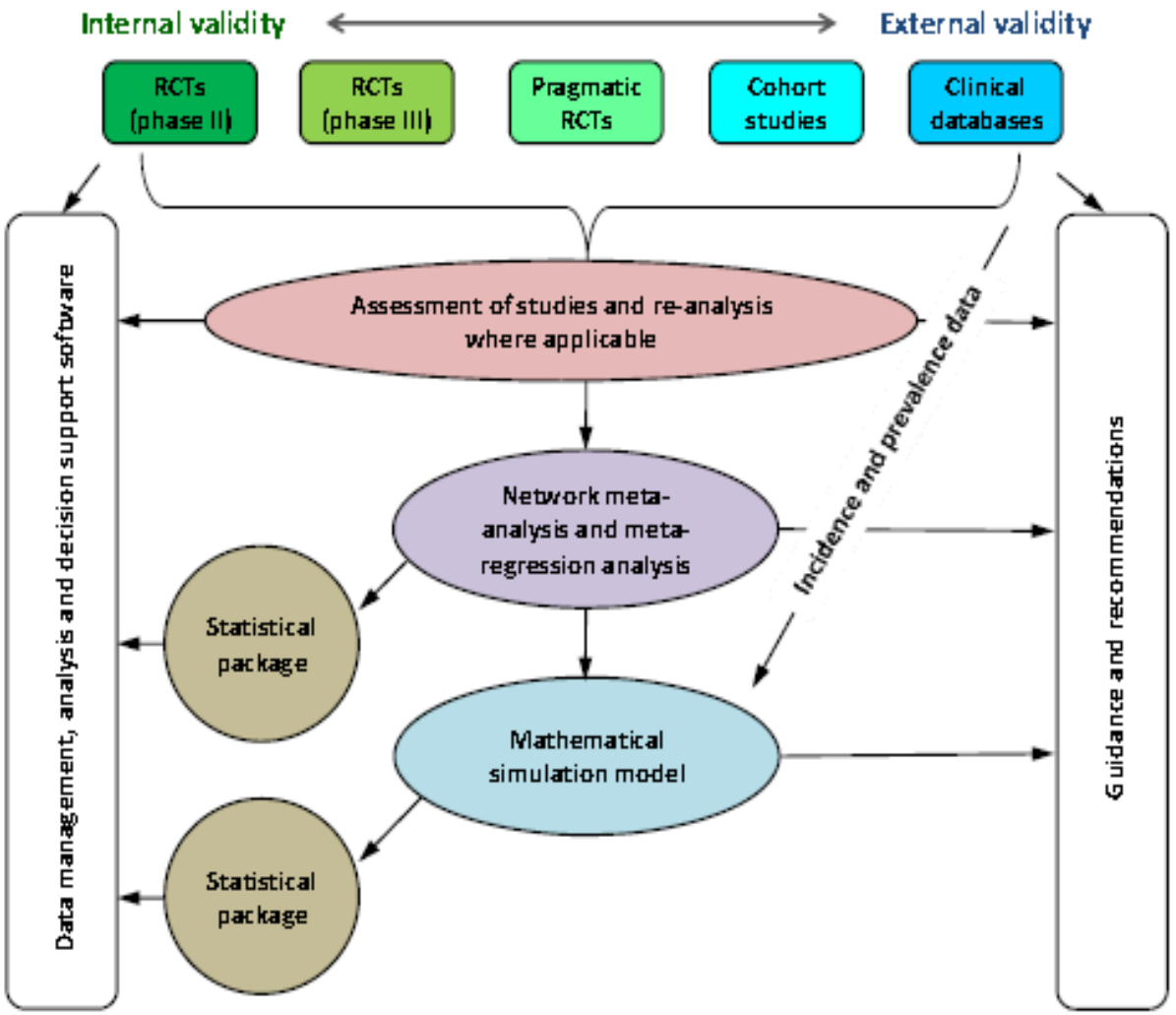
- **Frameworks developed jointly by Regulatory, HTA and Industry experts for use in:**
  - R&D strategy development, study design (comparators, endpoints, patients, care protocol)
  - Early Scientific Advice
  - HTA reviews of evidence base
- **Practical solutions: enable implementation of studies of greater value for RE assessment**
  - Translation from theory to practice
  - Regulatory and ethical reviews
  - Infrastructure and capability requirements / training & education
- **Advances in methodology to reliably predict effectiveness from available data**
  - Support extrapolation from optimised PIIIa studies
  - Increase acceptability of innovative PIIIb study data in evidence synthesis
  - Define the focus for post launch commitments
- **Aligning innovation in evidence generation with evolution of regulatory & HTA processes**
  - Understand how to evolve processes in a coordinated way without unnecessarily raising burden of evidence generation
  - Signal/avoid unintended consequences
  - Share insights and seek alignment with initiatives outside EU

Relationship between the WPs



Developing a predictive model for relative effectiveness (WP4)

Flow of work



Tasks

1) Identify suitable case-studies

2) Assess patient characteristics and risk of bias

3) Re-analyze individual patient data if available

4) Obtain best estimates of RE for different patient groups

5) Predict RE and absolute benefits and harms in different patient groups

6) Develop user-friendly software

7) Develop guidance and recommendations

STEPS INVOLVED IN AN NMA

Steps for conducting NMA

Step 1. Research project plan

- Objectives
- Populations
- Endpoints
- Comparators
- Any subgroups/sub-populations of interest
- Define systematic review (protocol)
- Analysis methods
- Limitations and biases

Workshop 1: NMA Challenges

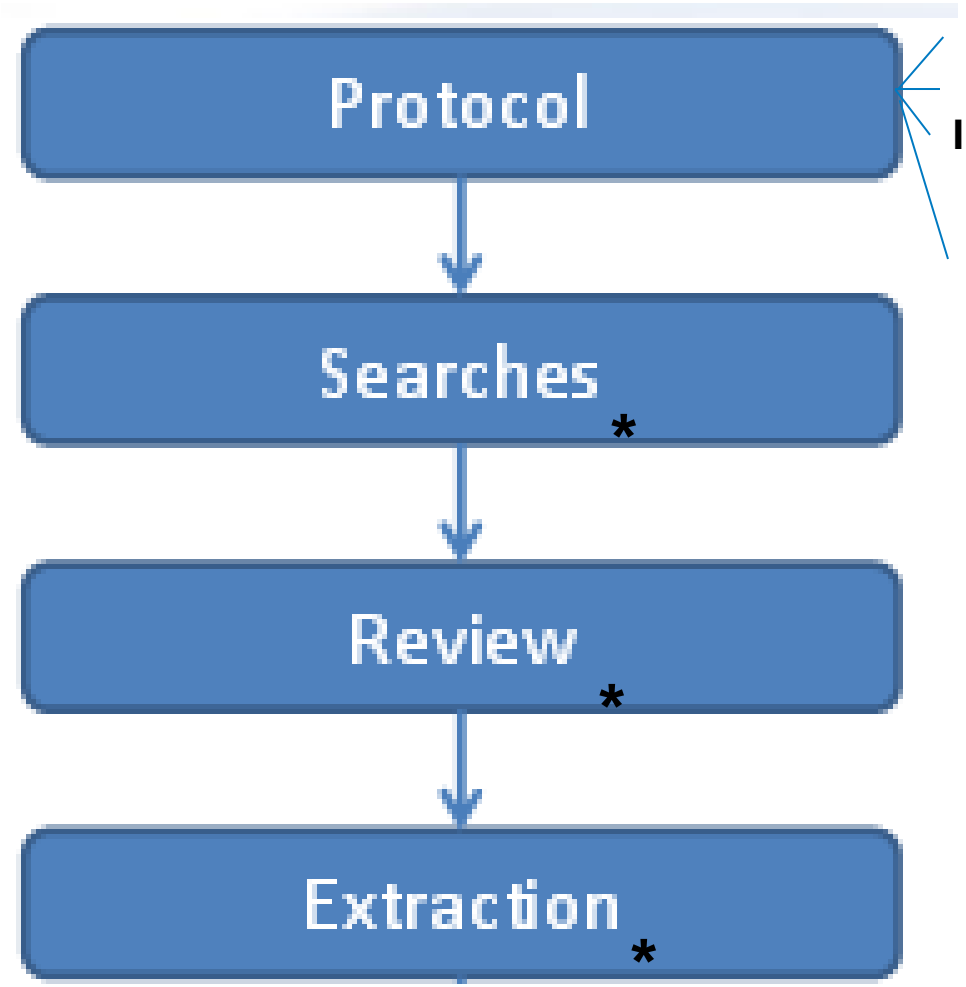
- What are the challenges in conducting a NMA?

Analysis methods: considering sources of heterogeneity

- Differences in inclusion/exclusion criteria or baseline characteristics
- Variability in control and treatment
  - Examples include dose, timing, brand
- Broader variability in management
  - Examples include care setting, co-medication, intermediate outcomes/crossovers, wash-in/out, compliance
- Differences in outcome measures
  - Examples include follow-up times, outcome definitions
- Variation in analysis
  - Examples include withdrawals, drop-outs, stopping rules, cross-overs
- Quality in design and execution, with bias or imprecision

NICE Technical Series Documents 2011: http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm

Step 2. Systematic review



Databases
Search criteria
Inclusion/exclusion criteria
- Study identification
Extraction and analysis

*** May be conducted by external vendor**

Step 3. Analysis – understanding evidence

- Understand scope of clinical package
- Critically assess the data
  - Clinical and statistical sources of heterogeneity
- How much direct (head to head) data is available?
- Define what “common comparators” exist
- What indirect comparisons/mixed treatment comparisons can be assessed?
- Develop network diagram / treatment comparison grid

Step 3. Analysis – conduct planned analysis

- Summarise direct head to head comparisons using meta-analysis
- Conduct NMA
- Investigate heterogeneity and inconsistent treatment effects (exchangeability assumption)
- Conduct meta-regression analyses to explore important prognostic variables and extensive sensitivity analyses
- Assess the statistical heterogeneity

Step 4. Reporting

- Summarise the evidence package
  - Sources of clinical and statistical heterogeneity
- Present summaries (tables, graphs) of head to head data, indirect comparisons and mixed treatment comparisons
- Provide interpretation of results
- Describe extent of heterogeneity
- Describe limitations and potential biases

Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparison Good Research Practices: Part 1 50

NMA CASE STUDY - IMMUNE THROMBOCYTOPENIA*

*** Funded by Amgen**

International Journal of Technology Assessment in Health Care, 28:3 (2012), 249–258.
© Cambridge University Press 2012
doi:10.1017/S0266462312000414

ROMIPLOSTIM AND ELTROMBOPAG FOR IMMUNE THROMBOCYTOPENIA: METHODS FOR INDIRECT COMPARISON

Katy L. Cooper
email: k.l.cooper@sheffield.ac.uk
Patrick Fitzgerald
University of Sheffield

Kerry Dillingham, Kawitha Helme
Amgen Limited
Ron Akehurst
University of Sheffield

Objectives: Immune thrombocytopenia (ITP) causes increased platelet destruction and suboptimal platelet production, increasing risk of bleeding. This analysis uses a Bayesian metaregression model to indirectly compare effectiveness of the thrombopoietin mimetics romiplostim and eltrombopag for increasing platelet counts, and contrasts the results with those of non-Bayesian approaches.
Methods: Ten databases were searched during 2010. Placebo-controlled trials of 24 weeks' duration were included. An indirect comparison was undertaken using Bayesian metaregression, which includes all trials in a single model. This was compared with previous analyses in which data for each intervention were combined using simple pooling, logistic regression or meta-analysis, followed by indirect comparison of pooled values using the Bucher method.
Results: Two trials of romiplostim and one of eltrombopag were included. The indirect evidence suggests romiplostim significantly improves overall platelet response compared with eltrombopag. Bayesian metaregression gave an odds ratio (OR) for eltrombopag versus romiplostim of 0.11 (95 percent credible interval 0.02–0.66); *p* values and Bayesian posterior probabilities ranged from 0.01 to 0.05 for all analyses. There was no significant difference in durable platelet response in any of the analyses, although the direction of effect favored romiplostim (OR = 0.15; 95 percent credible interval, 0.01–1.88); *p* values and Bayesian posterior probabilities ranged from 0.08 to 0.40 across analyses. Results were relatively consistent between analyses.
Conclusions: Bayesian metaregression generated similar results to other indirect comparison methods, and may be considered the most robust as it incorporates all data in a single model and accounts appropriately for parameter uncertainty.

Keywords: Idiopathic thrombocytopenic purpura, Romiplostim, Eltrombopag, Statistics as topic, Review, Systematic

Immune (idiopathic) thrombocytopenia (ITP) is an autoimmune condition characterized by increased platelet destruction and suboptimal platelet production, resulting in low platelet counts (thrombocytopenia) (21). Patients experience bleeding-related symptoms ranging from minor bruising to severe gastrointesti-

(21). Following splenectomy, approximately two-thirds of patients achieve sustained response for at least 5 years, with others having partial or transient responses. Approximately 14 percent do not respond, while 20 percent of responders later relapse. Complications of splenectomy include surgical morbidity and

Aims and objectives

- To validate the results of previous indirect comparisons of romiplostim and eltrombopag supporting HTAs (NICE)
- To explore additional statistical methods for the indirect comparison of romiplostim and eltrombopag:
  - In particular, determine statistical methods that allow more robust consideration of parameter uncertainty (that is, heterogeneity between studies) for indirect comparison

Methods

Systematic review to identify relevant RCTs

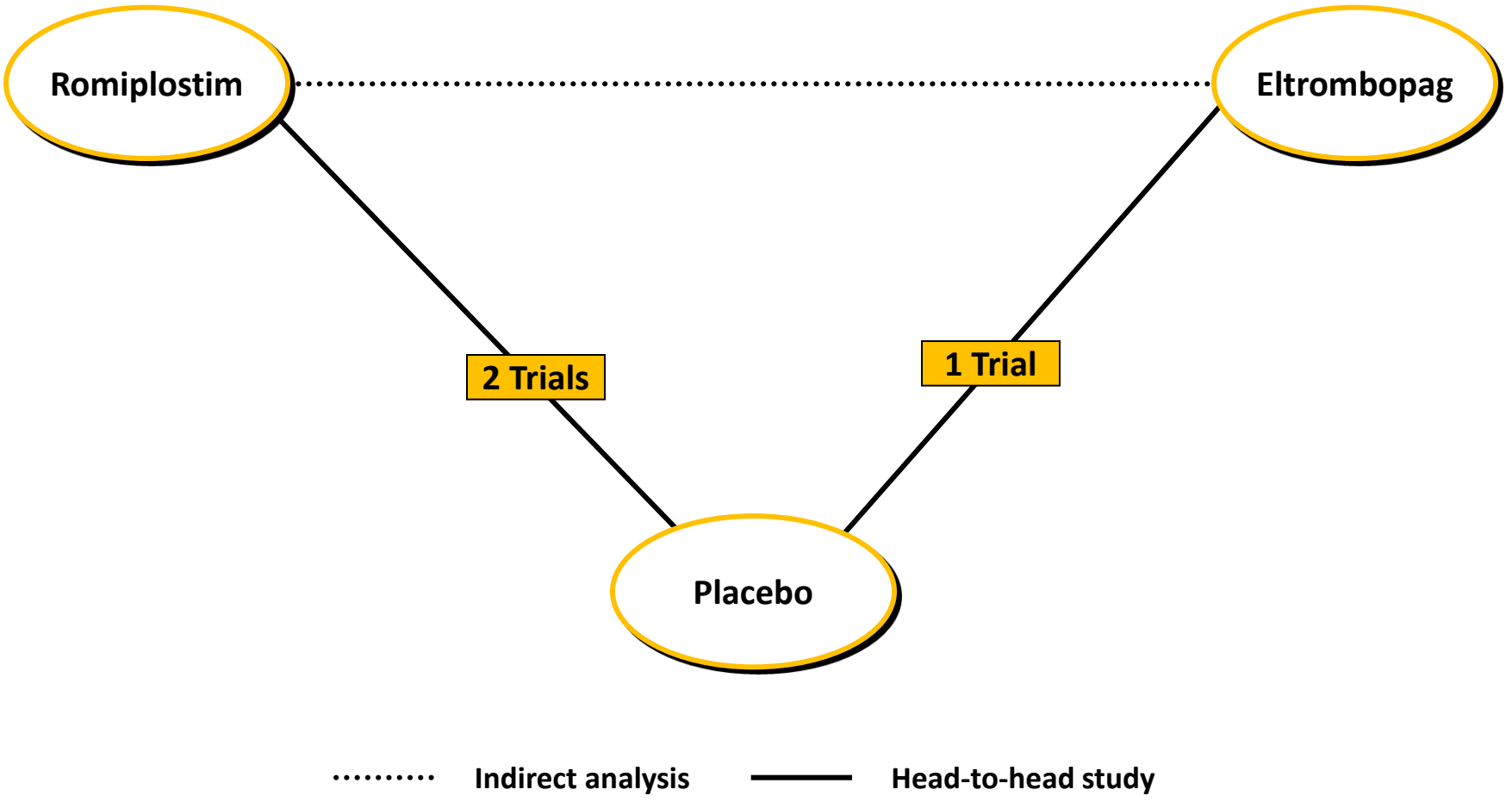
- **Inclusion criteria:**
- **RCTs comparing romiplostim or eltrombopag vs placebo for management of ITP**

Four RCTs each identified for romiplostim and eltrombopag

- **Inclusion criteria:**
- **Treatment duration ≥24 weeks and double-blind**
- **Reported data included platelet response**

Final selection: Two romiplostim RCTs and one eltrombopag RCT

Network diagram



Basis for Indirect Comparison

	Eltrombopag	Placebo	Romiplostim	Placebo
Overall response				
Non-splenectomised	51/85 (60%)	5/41 (12%)	36/41 (88%)	3/21 (14%)
Splenectomised	26/50 (52%)	2/21 (10%)	33/42 (79%)	0/21 (0%)
Overall	77/135 (57%)	7/62 (11%)	69/83 (83%)	3/42 (7%)
Durable response				
Non-splenectomised	38/85 (45%)	3/41 (7%)	25/41 (61%)	1/21 (5%)
Splenectomised	19/50 (38%)	1/21 (5%)	16/42 (38%)	0/21 (0%)
Overall	57/135 (42%)	4/62 (6%)	41/83 (49%)	1/42 (2%)

Definitions of Platelet Response Data Used for Indirect Comparison

	Eltrombopag	Romiplostim
Timing of definition for outcome measure	*Post hoc* analyses	*A priori*
Overall platelet response	Percentage of patients with a platelet count ≥50 and ≤400 x 10^9/L for ≥4 consecutive* weeks, excluding those receiving rescue medication during the assessment following a platelet response	Percentage of patients with a platelet count ≥50 x 10^9/L on ≥4 weeks during the trial, excluding responses within 8 weeks after rescue medications
Durable platelet response	Percentage of patients with platelet count ≥50 and ≤400 x 10^9/L on ≥6 of the last 8 weeks of treatment, excluding subjects who received rescue medication*	Percentage of patients with platelet count ≥50 x 10^9/L on ≥6 of the last 8 weeks of treatment, with no rescue medications at any time during the trial

Statistical Methods for Indirect Comparisons

Table 4. Indirect Comparison of Eltrombopag and Romiplostim

Analysis method
Previous analyses in STA submission and ERG report
Analysis 1 (eltrombopag STA): Summing of romiplostim data across trial arms then Bucher indirect comparison
Analysis 2 (ERG report): Pooling of romiplostim data via logistic regression (fixed treatment effects) then Bucher indirect comparison
Alternative methods for indirect comparison
Analysis 3: Meta-analysis of romiplostim data (Mantel-Haenszel weighting) then Bucher indirect comparison
Analysis 4: Pooling of romiplostim data via logistic regression (random treatment effects) then Bucher indirect comparison*
Analysis 5: Bayesian metaregression of romiplostim and eltrombopag data (random treatment effects)*

Bucher's Method

$$OR_{ab} = \frac{OR_{ap}}{OR_{bp}}$$

$$\log(OR_{ab}) = \log(OR_{ap}) - \log(OR_{bp})$$

Bucher HC, Guytt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J.Clin.Epi (1997) 60(6):683-91.

Bayesian metaregression

- $x_{ijk} \sim$ *binomial*(p_{ijk}, n_{ijk}), $\quad i = 1, 2, 3; j = 1, 2, 3; k = 1, 2$
- *logit*$(p_{ijk}) = \alpha_i + d_{i(2)}$ (if j = 2) + $d_{i(3)}$ (if j = 3) + β (if k = 2)
- $d_{i(2)} \sim N(\delta_2, \sigma_d^2)$, $i = 1, 2$; $d_{i(3)} \sim N(\delta_3, \sigma_d^2)$, $i = 3$; $\sigma_d^2 \sim$ *uniform*$(0, 0.6)$

Where

- x_{ijk} denotes the frequency of platelet response for each trial (i = 1, 2, 3), treatment group (j = 1 (placebo), 2 (romiplostim) or 3 (eltrombopag)) and splenectomy group (k = 1 (non-splenectomised), 2 (splenectomised)).
- $\alpha_i = \log\{p_{i1k}/(1 - p_{i1k})\}$ denote the fixed "study effect" (the log-odds of response for placebo-treated patients) in the *i*-th trial
- $d_{i(j)}$ denote the "treatment effect" (log OR for romiplostim or eltrombopag versus placebo) for each trial
- β denote the log OR for the effect of splenectomy, which is assumed to be common across all trials and treatment types

Bayesian metaregression (cont.)

- The model was used to estimate log OR for romiplostim versus placebo (δ_2) and for eltrombopag versus placebo (δ_3).

- The indirect log OR for eltrombopag versus romiplostim was then estimated from the posterior distribution of the difference $\delta_3 - \delta_2$.

Trial	Treatment arm	Splenectomy status	Logit model
Romiplostim (splenectomised)	placebo	splenectomised	$\alpha_1 + \beta$
	active	splenectomised	$\alpha_1 + \beta + d_{1(2)}$
Romiplostim (non-splenectomised)	placebo	non-splenectomised	α_2
	active	non-splenectomised	$\alpha_2 + d_{2(2)}$
Eltrombopag	placebo	splenectomised	$\alpha_3 + \beta$
	active	splenectomised	$\alpha_3 + \beta + d_{3(3)}$
	placebo	non-splenectomised	α_3
	active	non-splenectomised	$\alpha_3 + d_{3(3)}$

Overall Response – INCORRECT analysis

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	77/135 (57%)	7/62 (11%)	Kuter	33/42 (79%)	0/21 (0%)
			Kuter	36/41 (88%)	3/21 (14%)
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Unadjusted	**Eltrombopag**	**Romiplostim**
Overall Response	**77/135 (57%)**	**69/83 (83%)**
Odds Ratio		**OR=0.27**

Overall Response – Analysis 1

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	77/135 (57%)	7/62 (11%)	Kuter	33/42 (79%)	0/21 (0%)
			Kuter	36/41 (88%)	3/21 (14%)
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Adjusted	Eltrombopag	Romiplostim
Overall Response	OR=10.4	OR=64.1
Odds Ratio		OR=0.16

Overall Response – Analysis 2

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	77/135 (57%)	7/62 (11%)	Kuter	33/42 (79%)	0/21 (0%)
			Kuter	36/41 (88%)	3/21 (14%)
				Logistic Regression (Fixed)	
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Adjusted	**Eltrombopag**	**Romiplostim**
Overall Response	**OR=10.4**	**OR=77.7**
Odds Ratio		**OR=0.13**

Overall Response – Analysis 3

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	77/135 (57%)	7/62 (11%)	Kuter	33/42 (79%)	0/21 (0%)
			Kuter	36/41 (88%)	3/21 (14%)
				Meta-Analysis	
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Adjusted	**Eltrombopag**	**Romiplostim**
Overall Response	**OR=10.4**	**OR=68.4**
Odds Ratio		**OR=0.15**

Overall Response – Analysis 4

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	77/135 (57%)	7/62 (11%)	Kuter	33/42 (79%)	0/21 (0%)
			Kuter	36/41 (88%)	3/21 (14%)
				Logistic Regression (Random)	
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Adjusted	Eltrombopag	Romiplostim
Overall Response	OR=10.4	OR=105.8
Odds Ratio		OR=0.10

Overall Response – Analysis 5

	Eltrombopag	Placebo		Romiplostim	Placebo
Cheng	26/50 (52%)	2/21 (10%)	Kuter	33/42 (79%)	0/21 (0%)
	51/85 (60%)	5/41 (12%)	Kuter	36/41 (88%)	3/21 (14%)
			Bayesian Network		
Total	77/135 (57%)	7/62 (11%)		69/83 (83%)	3/42 (7%)

Adjusted	**Eltrombopag**	**Romiplostim**
Overall Response	**OR=11.6**	**OR=106.1**
Odds Ratio		**OR=0.11**

Cooper KL et al. (2012): Indirect Comparison Results for Overall Platelet Response

	OR eltrombopag vs placebo (95% CI)	OR romiplostim vs placebo (95% CI)	Indirect OR eltrombopag vs romiplostim (95% CI)
Analysis 1 (Eltrombopag STA)	10.4 (4.4, 24.6)	64.1 (17.3, 236.8)	0.16 (0.03, 0.78)
Analysis 2 (Eltrombopag ERG report)	10.4 (4.4, 24.6)	77.7 (19.5, 309.9)	0.13 (0.03, 0.68)
Analysis 3 (analysis 1 but with meta-analysis for pooling romiplostim data)	10.4 (4.4, 24.6)	68.4 (12.8, 365.6)	0.15 (0.02, 1.00)
Analysis 4 (analysis 2 but with random treatment effects and logistic regression for pooling romiplostim data)	10.4 (4.4, 24.6)	105.8 (24.6, 598.8)	0.10 (0.02, 0.57)
Analysis 5 (Bayesian meta-regression)	11.6 (4.4, 33.8)	106.1 (25.0, 593.5)	0.11 (0.02, 0.66)

Limitations of indirect comparisons

- Indirect comparisons are viewed as ‘observational evidence’
  - Trials may differ in patient population and trial design
- Differences in patient characteristics included:

Characteristic	Eltrombopag	Romiplostim
Required to have responded to first-line treatment	Yes	No
Splenectomised patients (%)	36%	50%
Patients receiving concomitant ITP medications at baseline (%)	Slightly higher	–
Patients having received ≥3 prior therapies (%)	–	Slightly higher
Patients withdrawing from the study (%)	Higher	–

Summary

- Consistent results were obtained across all of the statistical methods explored in this study
- The Bayesian metaregression approach generated similar results to other indirect comparison methods and may be considered the most robust of the analyses
  - It incorporates all trial data in a single model and accounts appropriately for parameter uncertainties

NMA METHODOLOGY

Sources for data.

- In the context of this course data values are extracted from published papers or internal company reports.
- They are summary Statistics, or more recently, estimated parameters from models.
  - i.e. Not primary data.

- Nearly always they will have been presented within a frequentist paradigm.
  - Estimates, standard errors and perhaps confidence intervals and “P-values”.

Extracting data

- Extracting data values from a paper is a time consuming job that requires skill.
  - Often the required piece of information is hidden in the text.
  - Tables in published papers are often reserved for the “best looking” analysis rather than the primary analysis.

- If all published papers followed the CONSORT Statement, then life would be much easier.

Ref: Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010, 11:32. (24 March 2010)

CONSORT statement

Item 17a –

For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).

Table 6- Example of reporting of summary results for each study group (Continuous outcomes)

(Adapted from table 3 of van Linschoten(234))

	Exercise therapy (n=65)			Control (n=66)		
	Baseline (mean (SD))	12 months (mean (SD))		Baseline (mean (SD))	12 months (mean (SD))	Adjusted difference* (95% CI) at 12 months
Function score (0-100)	64.4 (13.9)	83.2 (14.8)		65.9 (15.2)	79.8 (17.5)	4.52 (-0.73 to 9.76)
Pain at rest (0-100)	4.14 (2.3)	1.43 (2.2)		4.03 (2.3)	2.61 (2.9)	-1.29 (-2.16 to -0.42)
Pain on activity (0-100)	6.32 (2.2)	2.57 (2.9)		5.97 (2.3)	3.54 (3.38)	-1.19 (-2.22 to -0.16)

* Function score adjusted for baseline, age, and duration of symptoms.

Item 17b: Binary outcomes, Explanation
(my emphasis)

- When the primary outcome is binary, both the relative effect (risk ratio (relative risk) or odds ratio) and **the absolute effect (risk difference)** should be reported (with confidence intervals), as neither the relative measure nor the absolute measure alone gives a complete picture of the effect and its implications.

- Different audiences may prefer either relative or absolute risk, ...

Table 5 - Example of reporting of summary results for each study group (binary outcomes)*

(Adapted from table 2 of Mease et al(103))

Endpoint	Number (%)		Risk difference (95% CI)
	Etanercept (n=30)	Placebo (n=30)	
Primary endpoint			
Achieved PsARC at 12 weeks	26 (87)	7 (23)	63% (44 to 83)
Secondary endpoint			
Proportion of patients meeting ACR criteria:			
ACR20	22 (73)	4 (13)	60% (40 to 80)
ACR50	15 (50)	1 (3)	47% (28 to 66)
ACR70	4 (13)	0 (0)	13% (1 to 26)

Item 17b - For binary outcomes, **presentation of both absolute and relative effect sizes is recommended**

Example
"The risk of oxygen dependence or death was reduced by 16% (95% CI 25% to 7%). The absolute difference was -6.3% (95% CI -9.9% to -2.7%); early administration to an estimated 16 babies would therefore prevent 1 baby dying or being long-term dependent on oxygen" (also see table 7).(242)

Table 7 - Example of reporting both absolute and relative effect sizes
(Adpated from table 3 of The OSIRIS Collaborative Group(242))

outcome	Percentage (No)		Risk ratio (95% CI)	Risk difference (95% CI)
	Early administration (n=1344)	Delayed selective administration (n=1346)		
Death or oxygen dependence at "expected date of delivery"	31.9 (429)	38.2 (514)	0.84 (0.75 to 0.93)	-6.3 (-9.9 to -2.7)

- We will return to binary outcomes tomorrow.

Intention to Treat

- Usually we will want to extract ITT results from each study.

Intention to Treat (ITT) and Missing Data

- In a review of 403 RCTs published in 10 leading medical journals in 2002, 249 (62%) reported the use of intention-to-treat analysis for their primary analysis. This proportion was higher for journals adhering to the CONSORT statement (70% v 48%). **Among articles that reported the use of intention-to-treat analysis, only 39% actually analysed all participants as randomised, with more than 60% of articles having missing data in their primary analysis.(**221)
- Other studies show similar findings.(18) (222) (223) Trials with no reported exclusions are methodologically weaker in other respects than those that report on some excluded participants,(173) **strongly indicating that at least some researchers who have excluded participants do not report it.**

Ref: Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010, 11:32. (24 March 2010)

- We will return to the impact of missing data later.

Types of data.

- Usually we would compare treatment effects adjusted for baseline.
  - Like lsmeans and their SEs for off-time reduction.
  - Often these are not published and only raw means and standard errors are tabled.
- Sometimes we can reverse calculate SED from P values or confidence intervals.
- Two major forms of extraction:
  1. Absolute mean and SD for each treatment
     - Beware of distinction between SE for mean and SD.
  2. Differences with some measure of precision, often an SED or Confidence Interval.

Types of data.

- Often need a collection of different strategies. May include borrowing information across studies.

- Variability only known in some studies.
  - See Stevens (2013) for formal approach to doing this within Bayesian context.
  - This is only one of several possible Bayesian models.
  - Very easy in SAS 9.3 MCMC using new missing data methods.

[Stevens, J. 2013. *A note on dealing with missing standard errors in meta-analyses of continuous outcome measures in WinBUGS.* Pharmaceut. Statist. 2011, 10 374–378]

BMJ

RESEARCH

Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial

R van Linschoten, sports physician,[1] M van Middelkoop, researcher,[1] M Y Berger, researcher, general practitioner,[1] E M Heintjes, research associate,[4] J A N Verhaar, professor of orthopaedics,[2] S P Willemsen, statistician,[1,3] B W Koes, research professor,[1] S M Bierma-Zeinstra, associate research professor[1]

This is the original paper for the previous Normal data example.

Try extracting Function Score

Table 3 | Function and pain scores at 3 and 12 months follow-up

	Exercise therapy (n=65)			Control (n=66)				
	Baseline (mean (SD))	3 months (mean (SD))	12 months (mean (SD))	Baseline (mean (SD))	3 months (mean (SD))	12 months (mean (SD))	Adjusted difference* (95% CI) at 3 months	Adjusted difference* (95% CI) at 12 months
Function score (0-100)	64.4 (13.9)	78.8 (15.5)	83.2 (14.8)	65.9 (15.2)	74.9 (17.6)	79.8 (17.5)	4.92 (0.14 to 9.72)	4.52 (−0.73 to 9.76)
Pain at rest (0-10)	4.14 (2.3)	2.30 (2.5)	1.43 (2.2)	4.03 (2.3)	3.22 (2.8)	2.61 (2.9)	−1.07 (−1.92 to −0.22)	−1.29 (−2.16 to −0.42)
Pain on activity (0-10)	6.32 (2.2)	3.81 (2.9)	2.57 (2.9)	5.97 (2.3)	4.60 (3.0)	3.54 (3.38)	−1.00 (−1.91 to −0.08)	−1.19 (−2.22 to −0.16)

Mean scores are reported for those patients available at that time point. Adjusted differences are reported for the total available in analysis.
*Function score was adjusted for baseline score, age, and duration of symptoms. Pain at rest was adjusted for baseline score and age. Pain on activity was adjusted for baseline score, age, and gender. Positive adjusted differences for the function score, and negative difference for pain scores, are in favour of the exercise group.

We might extract the raw data means and ignore adjustments.

A: Mean= 83.2 SD=14.8 N=65

B: Mean= 79.8 SD=17.5 N=66

Difference = 3.40

Implied SED= $\sqrt{(14.8^2/65 + 17.5^2/66)} = 2.83$

Table 3 | Function and pain scores at 3 and 12 months follow-up

	Exercise therapy (n=65)			Control (n=66)				
	Baseline (mean (SD))	3 months (mean (SD))	12 months (mean (SD))	Baseline (mean (SD))	3 months (mean (SD))	12 months (mean (SD))	Adjusted difference* (95% CI) at 3 months	Adjusted difference* (95% CI) at 12 months
Function score (0-100)	64.4 (13.9)	78.8 (15.5)	83.2 (14.8)	65.9 (15.2)	74.9 (17.6)	79.8 (17.5)	4.92 (0.14 to 9.72)	4.52 (−0.73 to 9.76)
Pain at rest (0-10)	4.14 (2.3)	2.30 (2.5)	1.43 (2.2)	4.03 (2.3)	3.22 (2.8)	2.61 (2.9)	−1.07 (−1.92 to −0.22)	−1.29 (−2.16 to −0.42)
Pain on activity (0-10)	6.32 (2.2)	3.81 (2.9)	2.57 (2.9)	5.97 (2.3)	4.60 (3.0)	3.54 (3.38)	−1.00 (−1.91 to −0.08)	−1.19 (−2.22 to −0.16)

Mean scores are reported for those patients available at that time point. Adjusted differences are reported for the total available in analysis.
*Function score was adjusted for baseline score, age, and duration of symptoms. Pain at rest was adjusted for baseline score and age. Pain on activity was adjusted for baseline score, age, and gender. Positive adjusted differences for the function score, and negative difference for pain scores, are in favour of the exercise group.

Or use adjusted difference 4.52

Overall mean= (83.2*65 + 79.8*66)/131 = 81.49

Adjusted means

A: (81.49 + 4.52*66/131) = 83.77

B: (81.49 - 4.52*65/131) = 79.25

SED = (9.76 – (-0.73)) / (2 * 1.96) = 2.68 [2.65 if use T]

Smaller SED resulting from adjustment (was 2.83).

Effective SE for A is 2.68 * √(66/131) =1.90. SD for A=15.34

Effective SE for B is 2.68 * √(65/131) =1.89. SD for B=15.34

Using the P-value.

therapy on pain were 0.56 and 0.54, respectively. The difference in function scores at 12 months, however, did not reach statistical significance (4.52, 95% CI −0.73 to 9.76; P=0.09). The different between the two groups in the proportion of patients reporting "recovery" at 12 months was not significant.

- Or extract the SED from the P-value.

"The difference in function scores at 12 months, however, did not reach statistical significance (4.52, 95% CI −0.73 to 9.76; P=0.09)."

Assume equal two sided test,

SED= 4.52 / qnorm(1 - 0.045) = 2.67

Note how few digits are given here (could be P=0.08 or 0.10).

4.52 / qnorm(1 - 0.040)=2.58

4.52 / qnorm(1 - 0.050)=2.75

Data we hand into analysis

- The data we get from each trial may be either
  - Mean and SD or SE for each arm
  - Difference of means and SEDs.

- Analysis is easiest if form of data going into analysis is the same for every trial.
  - Mean and SD or SE for each arm
  - Difference of means and SEDs. Only need comparison to a single arm (control?) and not all comparisons.

CASE STUDY WITH NORMAL DATA

- This is a very small network, which we will use as an example.

- Usually networks are more extensive.

Example of network diagram

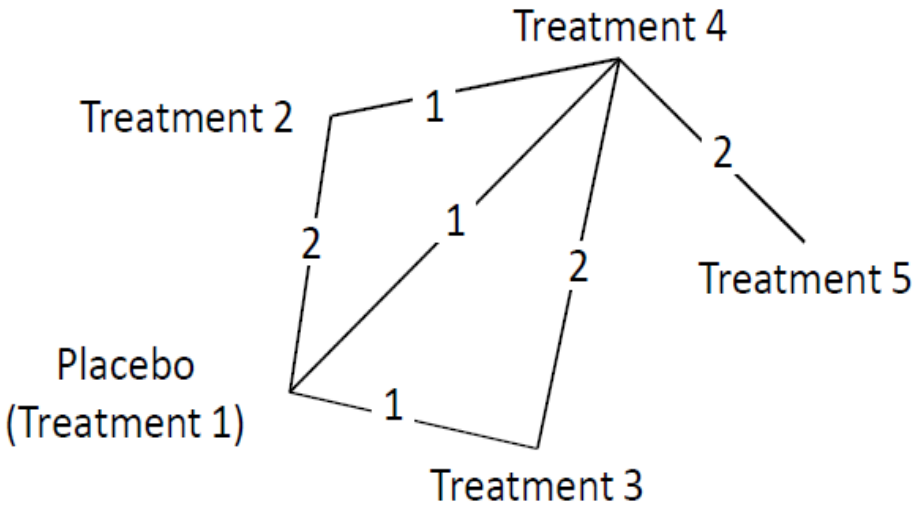


Figure 3 Parkinson network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison.

Parkinson's example

- **Mean "off-time" reduction in patients given dopamine agonists as adjunct therapy in Parkinson's disease.**

- **The available data are the mean, standard deviation and number of patients in each trial arm.**

- **Seven studies of five different drugs:**
  - **Placebo, coded 1,**
  - **Five active drugs coded 2 to 5.**

Example from NICE Decision Support Unit Technical Series Document

The Parkinson's data

Study	Treatment	y	sd	n	Difference [Calculated]	SE(diff) [Calculated]
1	1	-1.22	3.7	54		
	3	-1.53	4.28	95	-0.31	0.668
2	1	-0.7	3.7	172		
	2	-2.4	3.4	173	-1.7	0.383
3	1	-0.3	4.4	76		
	2	-2.6	4.3	71	-2.3	0.718
	4	-1.2	4.3	81	-0.9	0.695
4	3	-0.24	3	128		
	4	-0.59	3	72	-0.35	0.442
5	3	-0.73	3	80		
	4	-0.18	3	46	0.55	0.555
6	4	-2.2	2.31	137		
	5	-2.5	2.18	131	-0.3	0.274
7	4	-1.8	2.48	154		
	5	-2.1	2.99	143	-0.3	0.320

Differences from Standard analysis.

- Note how the SDs are estimated separately within treatment within Study and are not based on pooled variance within study.
  - This is quite common in this area when they are extracted from summary statistics rather than output from analysis.
  - Comparison of SDs across studies can be interesting.

Differences from Standard analysis.

- This example is not adjusted for other covariates
  - That is not adjusted for imbalance in covariate between arms within study.
  - Adjusted treatment differences can be used.
- Impact of changes in covariates between study is a different question.
  - That would be important if there is a Treatment by covariate interaction.
  - Meta-regression may be possible using summary for trial.
- Often based on Observed Cases, rather than an MAR analysis. Often handling of missing data is not mentioned in publications (though it should be).

Seen as an incomplete block design.

Study/Treatment	1	2	3	4	5
1	-1.22		-1.53		
2	-0.7	-2.4			
3	-0.3	-2.6		-1.2	
4			-0.24	-0.59	
5			-0.73	-0.18	
6				-2.2	-2.5
7				-1.8	-2.1

The Statistical model

- Simple two-way ANOVA.
- Study i and Arm k, with Treatment t(i,k)

$$Y_{ik} \sim N\left(\theta_{ik}, se_{ik}^2\right)$$

$$\theta_{ik} = \mu_i + \delta_{t(i,k)}$$

- Constraint that δ_j is zero for a chosen treatment, usually $\delta_1 = 0$.
- The notation δ is used here as it represents the *difference* from some reference treatment such as placebo, which may not be observed in this i'th trial.

Useful reference

- This paper is ideal ammunition for the Statistician in explaining how much of indirect comparisons comes down to this very simple two-way ANOVA model.

Piepho H. P., Williams E. R., and Madden L. V.. 2012. *The Use of Two-Way Linear Mixed Models in Multitreatment Meta-Analysis.* Biometrics.

DOI: 10.1111/j.1541-0420.2012.01786.x

And they are agricultural Statisticians!

All arms and all studies equally precise.

Not the recommended analysis as not efficient.
But it is unbiased.

Two- way ANOVA

```
proc mixed data=Parkinsons;
class Study Treatment;
model Y= Study Treatment /solution outp=Pred;
id SE Study Treatment;
lsmeans Treatment / diff=control("1");
run;
```

```
data Parkinsons;
input Study Treatment Y SD N ;
Var=SD*SD/N;
SE=sqrt(Var);
Weight=1/Var;
Record=_N_;
datalines;
1  1  -1.22    3.7     54
1  3  -1.53    4.28    95
2  1   -0.7    3.7    172
2  2   -2.4    3.4    173
3  1   -0.3    4.4     76
3  2   -2.6    4.3     71
3  4   -1.2    4.3     81
4  3   -0.24   3      128
4  4   -0.59   3       72
5  3   -0.73   3       80
5  4   -0.18   3       46
6  4   -2.2    2.31   137
6  5   -2.5    2.18   131
7  4   -1.8    2.48   154
7  5   -2.1    2.99   143
;
run;
```

The estimates. Note the d.f.

Effect	Study	Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			-2.1000	0.2576	4	-8.15	0.0012
Study	1		0.1426	0.4108	4	0.35	0.7460
Study	2		0.6705	0.4488	4	1.49	0.2094
Study	3		0.7137	0.3691	4	1.93	0.1253
Study	4		1.3782	0.3578	4	3.85	0.0183
Study	5		1.3382	0.3578	4	3.74	0.0201
Study	6		-0.4000	0.2974	4	-1.34	0.2499
Study	7		0	.	.	.	.
Treatment		1	0.8511	0.4316	4	1.97	0.1199
Treatment		2	-1.0921	0.4628	4	-2.36	0.0777
Treatment		3	0.3137	0.3979	4	0.79	0.4746
Treatment		4	0.3000	0.2974	4	1.01	0.3702
Treatment		5	0	.	.	.	.

Differences of Least Squares Means

Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	2	1	-1.9432	0.2895	4	-6.71	0.0026
Treatment	3	1	-0.5374	0.3201	4	-1.68	0.1685
Treatment	4	1	-0.5511	0.3127	4	-1.76	0.1528
Treatment	5	1	-0.8511	0.4316	4	-1.97	0.1199

Fixed effect analysis.

```
proc mixed data=Parkinsons;
class Study Treatment;
model Y= Study Treatment / ddf= 500, 500;
weight Weight;
parms 1 / hold=(1);
lsmeans Treatment / diff=control("1") df=500;
run;
```

Note no covariance parameters are estimated.

Variability is assumed known and fixed (so set denominator d.f. at large value).

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Study	6	500	6.14	<.0001
Treatment	4	500	7.94	<.0001

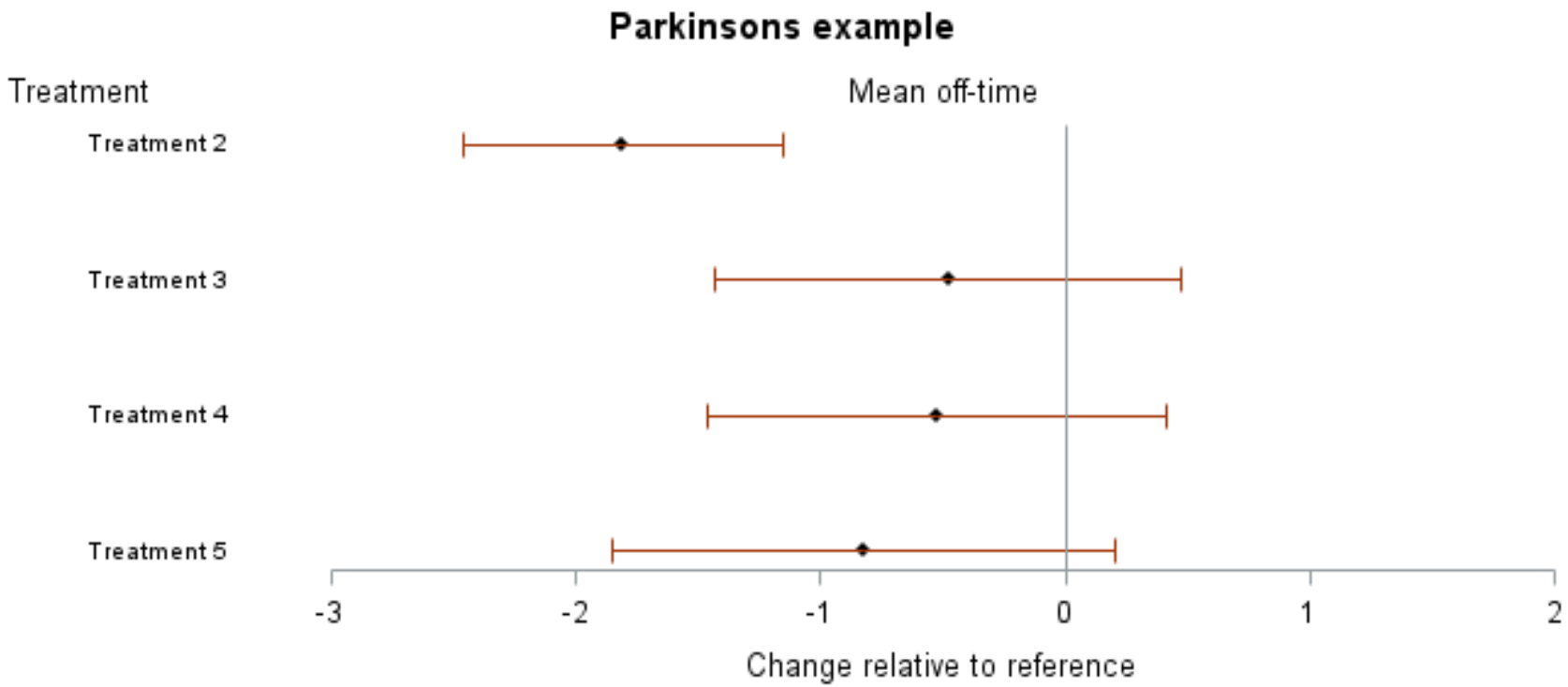
Least Squares Means

Effect	Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	1	-0.7445	0.3443	500	-2.16	0.0311
Treatment	2	-2.5561	0.3981	500	-6.42	<.0001
Treatment	3	-1.2226	0.2546	500	-4.80	<.0001
Treatment	4	-1.2685	0.2056	500	-6.17	<.0001
Treatment	5	-1.5685	0.2727	500	-5.75	<.0001

Differences of Least Squares Means

Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	2	1	-1.8116	0.3327	500	-5.45	<.0001
Treatment	3	1	-0.4781	0.4866	500	-0.98	0.3263
Treatment	4	1	-0.5240	0.4786	500	-1.09	0.2741
Treatment	5	1	-0.8240	0.5220	500	-1.58	0.1151

Mean and 95% CI for difference from Treatment 1.



Is this correct?

- We could rebuild the original data (subject data within each trial) and analyze that.

- Within each arm place data at either

(Mean + Delta) or (Mean - Delta)

where delta = SD $\sqrt{((n-1)/n)}$.

```
* Generate full data;
data Full;
set Parkinsons;
keep FullRec Study Treatment Response;
retain FullRec 0;
drop i;
M=N;
* Handle case of N being odd;
if mod(n,2) then do;
    Response=Y;
    FullRec=FullRec+1;
    output;
    M=M-1;
end;
Delta=SD*sqrt((N-1)/M);
do i=1 to M;
    Response=Y+Delta;
    FullRec=FullRec+1;
    output;
    Delta=-Delta;
end;
run;
```

Tabulate mean and SD as check.

Response										
	Treatment									
	1		2		3		4		5	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Study										
1	-1.22	3.70	.	.	-1.53	4.28	.	.	.	.
2	-0.70	3.70	-2.40	3.40	.	.	.	.	.	.
3	-0.30	4.40	-2.60	4.30	.	.	-1.20	4.30	.	.
4	.	.	.	.	-0.24	3.00	-0.59	3.00	.	.
5	.	.	.	.	-0.73	3.00	-0.18	3.00	.	.
6	.	.	.	.	.	.	-2.20	2.31	-2.50	2.18
7	.	.	.	.	.	.	-1.80	2.48	-2.10	2.99

Analysis model

```
proc mixed data=Full;
class Study Treatment;
model Response=Study Treatment /ddfm=kr;
lsmeans Treatment / diff=control("1");
repeated /subject=FullRec group=Study*Treatment;
run;
```

Note the unusual use of separate variances.

Individual variance for each Study*Arm

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Residual	FullRec	Study*Treatment 1 1	13.6307
Residual	FullRec	Study*Treatment 1 3	18.2842
Residual	FullRec	Study*Treatment 2 1	13.6831
Residual	FullRec	Study*Treatment 2 2	11.5551
Residual	FullRec	Study*Treatment 3 1	19.3198
Residual	FullRec	Study*Treatment 3 2	18.4157
Residual	FullRec	Study*Treatment 3 4	18.4421
Residual	FullRec	Study*Treatment 4 3	9.0001
Residual	FullRec	Study*Treatment 4 4	9.0004
Residual	FullRec	Study*Treatment 5 3	9.0203
Residual	FullRec	Study*Treatment 5 4	9.0626
Residual	FullRec	Study*Treatment 6 4	5.3275
Residual	FullRec	Study*Treatment 6 5	4.7449
Residual	FullRec	Study*Treatment 7 4	6.1414
Residual	FullRec	Study*Treatment 7 5	8.9180

Effectively recovered just the SD^2 for each combination.

Very nearly the same

Differences of Least Squares Means

Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	2	1	-1.8118	0.3338	490	-5.43	<.0001
Treatment	3	1	-0.4774	0.4899	334	-0.97	0.3305
Treatment	4	1	-0.5244	0.4819	364	-1.09	0.2772
Treatment	5	1	-0.8244	0.5252	503	-1.57	0.1171

Back to using summary data. What happens if I forget to use HOLD?

```
proc mixed data=Parkinsons ;
class Study Treatment;
model Y= Study Treatment / ddfm=kr;
weight weight;
lsmeans treatment / diff=control("1");
run;
```

Standard errors are too small.

Differences of Least Squares Means

Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	2	1	-1.8116	0.2516	4	-7.20	0.0020
Treatment	3	1	-0.4781	0.3680	4	-1.30	0.2637
Treatment	4	1	-0.5240	0.3620	4	-1.45	0.2213
Treatment	5	1	-0.8240	0.3948	4	-2.09	0.1052

- **Variability has been assessed from the between arm and study. So is too small here.**
- **Often the standard errors will be too large.**
- **Code looks sensible so beware!**

Using GENMOD for the correct analysis

```
proc genmod data=Parkinsons;
class Study Treatment;
model Y= Study Treatment / dist=normal noscale;
weight Weight;
Lsmeans Treatment / diff=control("1");
run;
```

- The NOSCALE option means that a scale parameter (residual) is not estimated but fixed at 1.

Same results using GENMOD.

Differences of Treatment Least Squares Means

Treatment	_Treatment	Estimate	Standard Error	z Value	Pr > |z|
2	1	-1.8116	0.3327	-5.45	<.0001
3	1	-0.4781	0.4866	-0.98	0.3259
4	1	-0.5240	0.4786	-1.09	0.2736
5	1	-0.8240	0.5220	-1.58	0.1144

Note on use of GENMOD

- GENMOD uses maximum likelihood rather than REML.
  - Not an issue here as we are not estimating the residual.

Summary
[Normal data: Fixed effects – Frequentist]

- Use summary values and assume variances fixed and known (do not estimate any covariance paramaters).
  - Use WEIGHT.
  - Use PARMS and HOLD with MIXED or GLIMMIX, or NOSCALE with GENMOD.

The dark side!

Going Bayesian ...

Why Bayesian?

- Today we will develop Bayesian solutions for the Normal case.

- Three reasons
  1. Many of the methods being promulgated, especially for Binary data are Bayesian.
  2. It provides a way to handle the heterogeneity when it is not well estimated from within the meta-analysis.
  3. It provides a way to fit complex hierarchical models which have been difficult to fit within a maximum likelihood paradigm (see tomorrow).

The NICE results (Bayesian)

Flat conjugate priors. Bound to be the same for Fixed effects!

Table A9 Parkinson example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Treatments 2 to 5 (d_{12} to d_{15}) relative to Placebo, absolute effects of Placebo (T_1) and treatments 2 to 5 (T_2 to T_5), heterogeneity parameter τ and model fit statistics for different data types.

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
	Arm-level data: Example 5							
d_{12}	-1.81	0.33	-1.81	(-2.46,-1.16)	-1.85	0.54	-1.84	(-2.91,-0.85)
d_{13}	-0.47	0.49	-0.47	(-1.43,0.49)	-0.50	0.66	-0.50	(-1.78,0.75)
d_{14}	-0.52	0.48	-0.52	(-1.46,0.43)	-0.53	0.65	-0.53	(-1.77,0.71)
d_{15}	-0.82	0.52	-0.82	(-1.84,0.22)	-0.83	0.80	-0.83	(-2.35,0.69)
T_1	-0.73	0.22	-0.73	(-1.16,-0.30)	-0.73	0.22	-0.73	(-1.16,-0.30)
T_2	-2.54	0.40	-2.54	(-3.32,-1.76)	-2.58	0.58	-2.57	(-3.72,-1.50)
T_3	-1.21	0.53	-1.20	(-2.25,-0.15)	-1.23	0.70	-1.23	(-2.57,0.10)
T_4	-1.25	0.53	-1.25	(-2.28,-0.21)	-1.26	0.69	-1.26	(-2.57,0.05)
T_5	-1.55	0.57	-1.55	(-2.66,-0.43)	-1.57	0.83	-1.56	(-3.14,0.02)
τ	-	-	-	-	0.40	0.43	0.28	(0.01,1.55)

NICE Winbugs code (Fixed effects model).

```
# Normal likelihood, identity link
# Fixed effects model for multi-arm trials
model{
for(i in 1:ns){
    mu[i] ~ dnorm(0,.0001) # vague priors for all trial baselines
    for (k in 1:na[i]) {
        var[i,k] <- pow(se[i,k],2)                                  # calculate variances
        prec[i,k] <- 1/var[i,k]                                     # set precisions
        y[i,k] ~ dnorm(theta[i,k],prec[i,k])                        # normal likelihood
        theta[i,k] <- mu[i] + d[t[i,k]] - d[t[i,1]] ← Unecessary complication   # model for linear predictor
        dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]           #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na[i]]) # summed residual deviance contribution for this trial
}
totresdev <- sum(resdev[])                                          #Total Residual Deviance
d[1]<-0                                                             # treatment effect is zero for reference treatment
for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }                            # vague priors for treatment effects
}
```

Easy Bayes in SAS.
Same GENMOD as before.

```
proc genmod data=Parkinsons;
class Study Treatment;
model Y= Study Treatment / dist=normal noscale;
bayes seed=1352 STATS(alpha=0.05 percent=2.5 25 50
    75 97.5 )=all;
weight Weight;
lsmeans Treatment / diff=control("1");
run;.
```

Mean and SD of posterior are very similar to ML.
Based on MCMC sample.
Theory says they are the same for one specific prior.

Sample Differences of Treatment Least Squares Means

Treatment	_Treatment	N	Estimate	Standard Deviation	Percentiles 2.5th	25th	50th	75th	97.5th
2	1	10000	-1.8150	0.3324	-2.4513	-2.0403	-1.8141	-1.5887	-1.1645
3	1	10000	-0.4867	0.4900	-1.4420	-0.8164	-0.4862	-0.1600	0.4780
4	1	10000	-0.5302	0.4798	-1.4703	-0.8534	-0.5367	-0.2028	0.4055
5	1	10000	-0.8302	0.5220	-1.8638	-1.1793	-0.8337	-0.4734	0.1946

These are summaries of the sampled posterior distribution.

	FE model mean	sd	median	CrI
				Arm-level d
d_{12}	-1.81	0.33	-1.81	(-2.46,-1.16)
d_{13}	-0.47	0.49	-0.47	(-1.43,0.49)
d_{14}	-0.52	0.48	-0.52	(-1.46,0.43)
d_{15}	-0.82	0.52	-0.82	(-1.84,0.22)

Markov Chain Monte Carlo (MCMC)
Sample from the posterior.

- When we use MCMC to solve a Bayesian problem we get a sample from the posterior distribution.
  - It is only a sample, so run it again (new seed) and you get a different sample.
  - Need to worry about Markov Chain error (accuracy of our statistics).
  - The sample is usually autocorrelated.

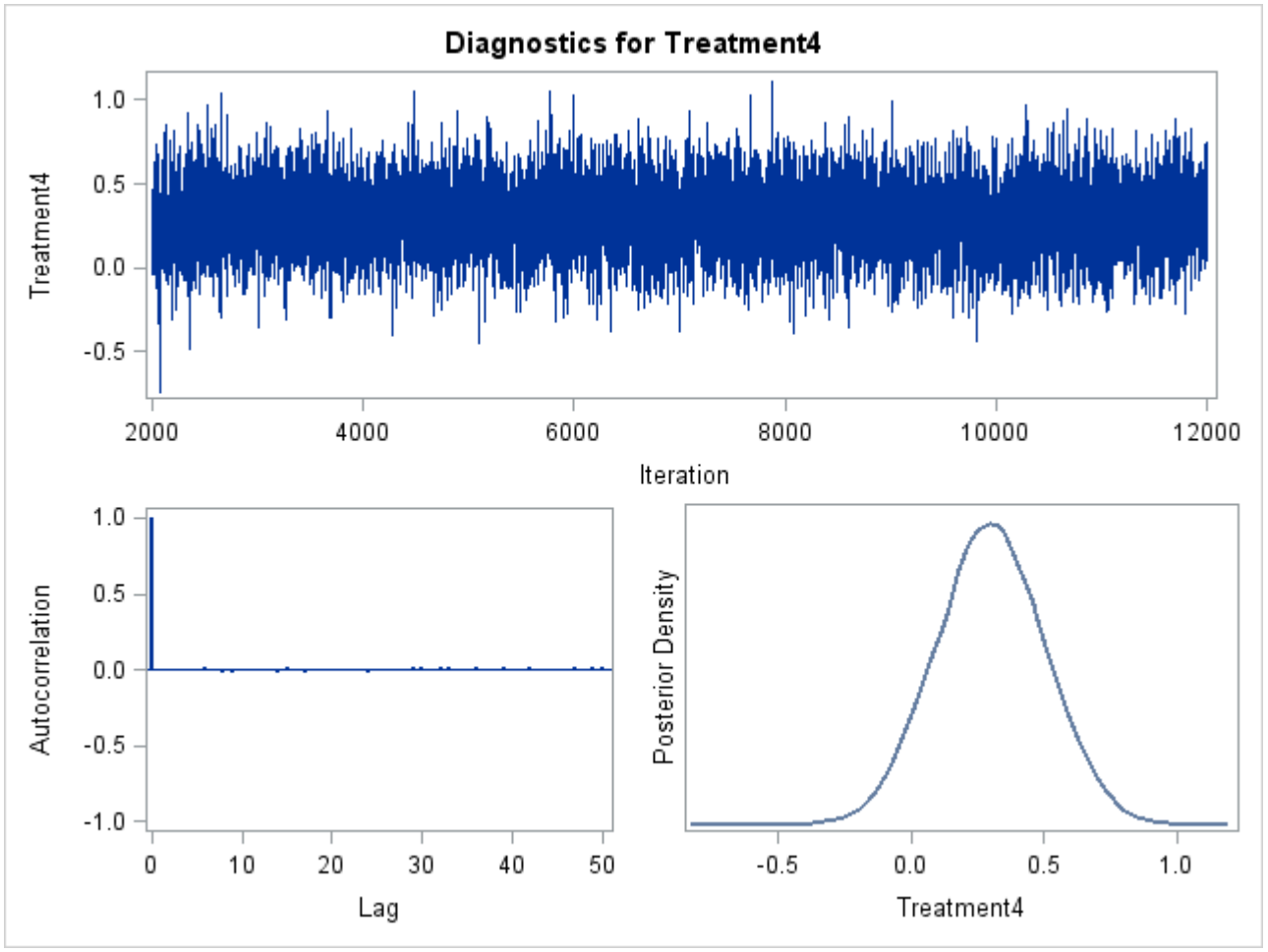
- We estimate the real properties of the posterior distribution from the sample.
  - Sample mean for mean of posterior.
  - Sample percentiles estimate percentiles for posterior distribution, such as median.

Sample from the posterior.

- For any **statistic derived from the model parameters**, we calculate the value for each member of the sample, and we have a sample from its posterior.
  - Odds ratio from parameters in logistic model.
  - Log(HR) from Hazard Ratio.

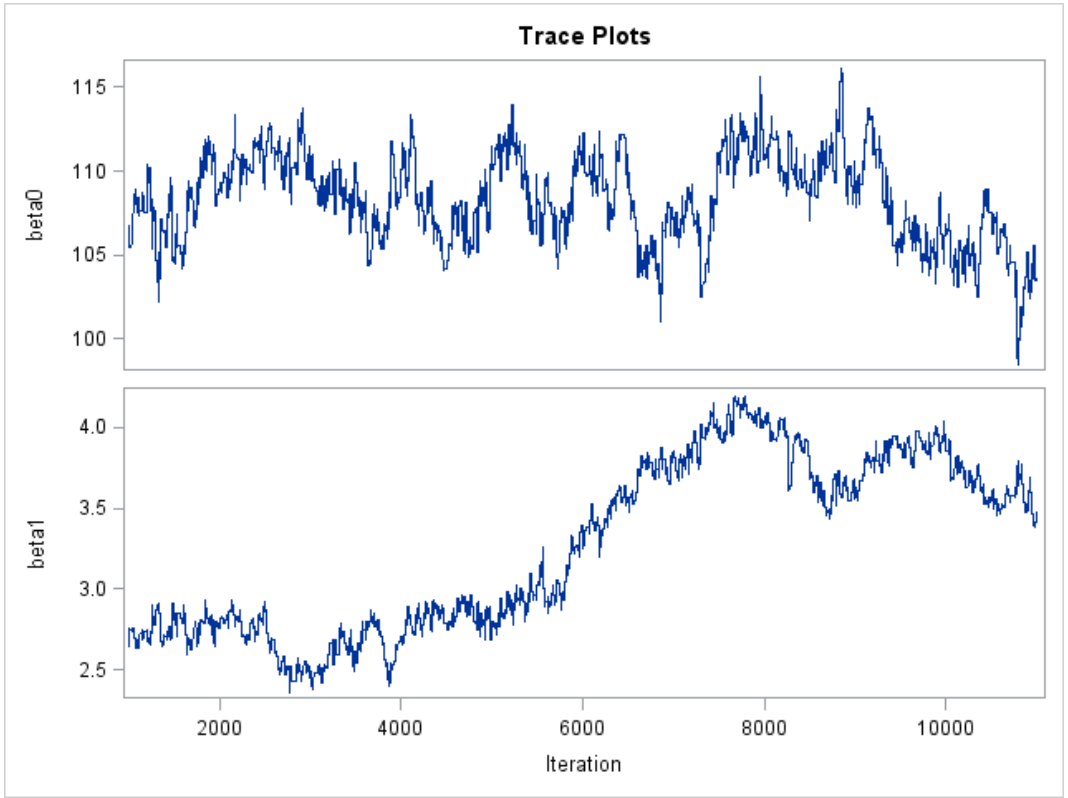
- Important to realize that the Markov chain is stepping around in the parameter space, and the frequency of times it chooses a point is proportional to the posterior probability.
  - Unlike Winbugs the Metropolis-Hastings algorithm in the MCMC procedure can repeat a point in the parameter space.

GENMOD can spot Conjugancy.
This allows direct sampling.



Understanding the Diagnostic plots

- Trace Plots
  - Is the chain stationary and mixing?
  - Constant mean, constant variance.
  - Moving around the parameter space freely.
  - Moving rapidly between extremes.



What you do not want!

Diagnostic plots

- Autocorrelation
  - Measures the correlation between each draw and its kth lag.
  - The further the lag from the original measure the smaller you expect the correlation to be.
  - High correlation between distant draws suggests poor mixing.

Diagnostic plots

- Density estimate
  - This is a Kernel density estimate, and will behave badly at a boundary.
  - So use with care for Variances and SDs., especially variance components, where likelihood may be increasing at the boundary.

    [Later we show how to use SGPLOT to get a better picture.]

The MCMC procedure in SAS.

- Here we introduce the MCMC procedure in SAS.

- It is important which version of SAS you are using.
  - SAS 9.2 Make sure you are using Level 2M3.
  - SAS 9.3 Has many new features that make coding easier, and the procedure run faster.

- It basically does the things that Winbugs does, but in a slightly different way.

More difficult Bayesian solution in SAS. The MCMC procedure. (SAS 9.2)

```
ods graphics on;
proc mcmc data=Parkinsons ntu=1000 nmc=200000  thin=20 seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
parms P_Study1-P_Study7 0;
parms P_Treat2-P_Treat5 0;
prior P_Study1-P_Study7 ~ general(0);
prior P_Treat2-P_Treat5 ~ general(0);
p_Treat[1]=0;
mu=  P_Study[Study] + P_Treat[Treatment] ;
model Y ~ normal(mean=Mu, sd=SE);
run;
```

- Note the similarity to NLMIXED code. / See next slide(s)

Declare the fixed effects parameters and constraint

```
ods graphics on;
proc mcmc data=Parkinsons ntu=1000 nmc=200000  thin=20 seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
parms P_Study1-P_Study7 0;
parms P_Treat2-P_Treat5 0;
prior P_Study1-P_Study7 ~ general(0);
prior P_Treat2-P_Treat5 ~ general(0);
p_Treat[1]=0;
mu=  P_Study[Study] + P_Treat[Treatment] ;
model Y ~ normal(mean=Mu, sd=SE);
run;
```

- Set fixed effects constraint with treat effect for treatment 1 as zero.

Set priors

```
ods graphics on;
proc mcmc data=Parkinsons ntu=1000 nmc=200000  thin=20 seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
parms P_Study1-P_Study7 0;
parms P_Treat2-P_Treat5 0;
prior P_Study1-P_Study7 ~ general(0);
prior P_Treat2-P_Treat5 ~ general(0);
p_Treat[1]=0;
mu=  P_Study[Study] + P_Treat[Treatment] ;
model Y ~ normal(mean=Mu, sd=SE);
run;
```

- general(0) is a completely flat (improper) prior.

Declare the model

```
ods graphics on;
proc mcmc data=Parkinsons ntu=1000 nmc=200000  thin=20  seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
parms P_Study1-P_Study7 0  P_Treat2-P_Treat5 0;
prior P_Study1-P_Study7 ~ general(0);
prior P_Treat2-P_Treat5 ~ general(0);
P_Treat[1]=0;
Mu=  P_Study[Study] + P_Treat[Treatment] ;
model Y ~ normal(mean=Mu, sd=SE);
run;
```

10,000 with a thin of 20 is not enough ... if you want to report to 2 decimal places

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
P_Study1	10000	-1.1245	0.4159	-1.4084	-1.1312	-0.8414
P_Study2	10000	-0.6371	0.2662	-0.8143	-0.6393	-0.4578
P_Study3	10000	-0.5921	0.3530	-0.8336	-0.5965	-0.3536
P_Study4	10000	0.1083	0.4840	-0.2208	0.0973	0.4330
P_Study5	10000	-0.0563	0.5104	-0.4146	-0.0583	0.2862
P_Study6	10000	-1.7022	0.4784	-2.0333	-1.7093	-1.3812
P_Study7	10000	-1.3051	0.4808	-1.6367	-1.3090	-0.9880
P_Treat2	10000	-1.8144	0.3427	-2.0448	-1.8149	-1.5870
P_Treat3	10000	-0.4607	0.4661	-0.7695	-0.4526	-0.1419
P_Treat4	10000	-0.4980	0.4491	-0.7965	-0.4916	-0.1933
P_Treat5	10000	-0.7965	0.4908	-1.1301	-0.7895	-0.4592

Monte Carlo Standard Errors

Parameter	MCSE	Standard Deviation	MCSE/SD
P_Treat2	0.00946	0.3427	0.0276
P_Treat3	0.0374	0.4661	0.0803
P_Treat4	0.0335	0.4491	0.0745
P_Treat5	0.0351	0.4908	0.0716

100,000 with a thin of 20...

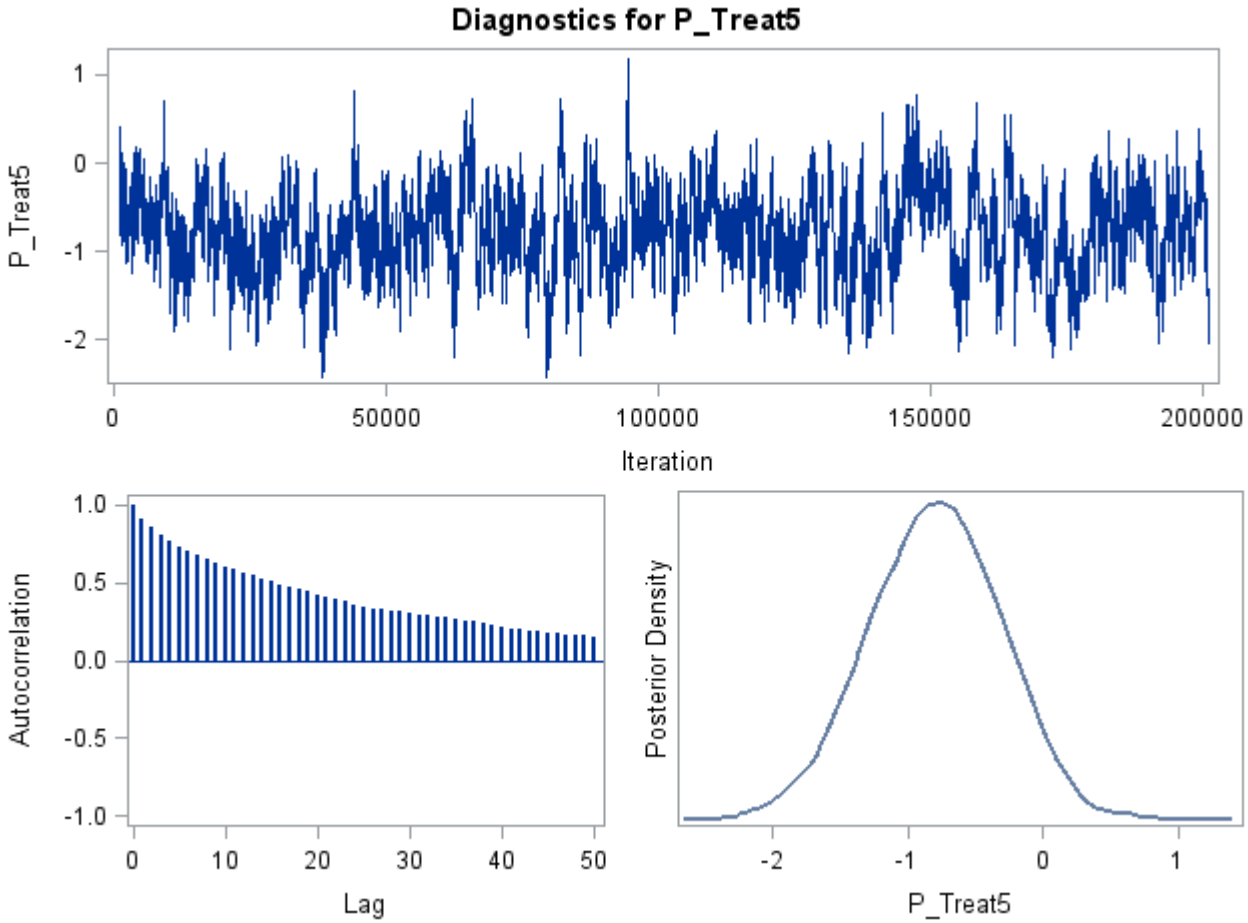
Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
P_Study1	100000	-1.1335	0.4335	-1.4275	-1.1326	-0.8424
P_Study2	100000	-0.6411	0.2612	-0.8166	-0.6404	-0.4646
P_Study3	100000	-0.5967	0.3652	-0.8439	-0.5969	-0.3504
P_Study4	100000	0.1119	0.5107	-0.2331	0.1155	0.4615
P_Study5	100000	-0.0510	0.5375	-0.4150	-0.0484	0.3170
P_Study6	100000	-1.6947	0.5178	-2.0404	-1.6923	-1.3453
P_Study7	100000	-1.2941	0.5182	-1.6401	-1.2922	-0.9446
P_Treat2	100000	-1.8081	0.3324	-2.0325	-1.8078	-1.5846
P_Treat3	100000	-0.4628	0.4924	-0.8006	-0.4662	-0.1291
P_Treat4	100000	-0.5068	0.4866	-0.8377	-0.5085	-0.1814
P_Treat5	100000	-0.8056	0.5313	-1.1658	-0.8078	-0.4506

Monte Carlo Standard Errors

Parameter	MCSE	Standard Deviation	MCSE/SD
P_Treat2	0.00241	0.3324	0.00725
P_Treat3	0.00927	0.4924	0.0188
P_Treat4	0.0100	0.4866	0.0206
P_Treat5	0.0106	0.5313	0.0200

Treatment 5 – Treatment 1



Other ways to improve the MCSE

- Different method for the proposal distribution.
  - propcov=quanew on MCMC statement.
- Modify the arrangement of parameters into blocks using the PARMS statements.
- Idea is to reduce the autocorrelation.

10,000 only, using propcov=quanew and a single PARMS statement

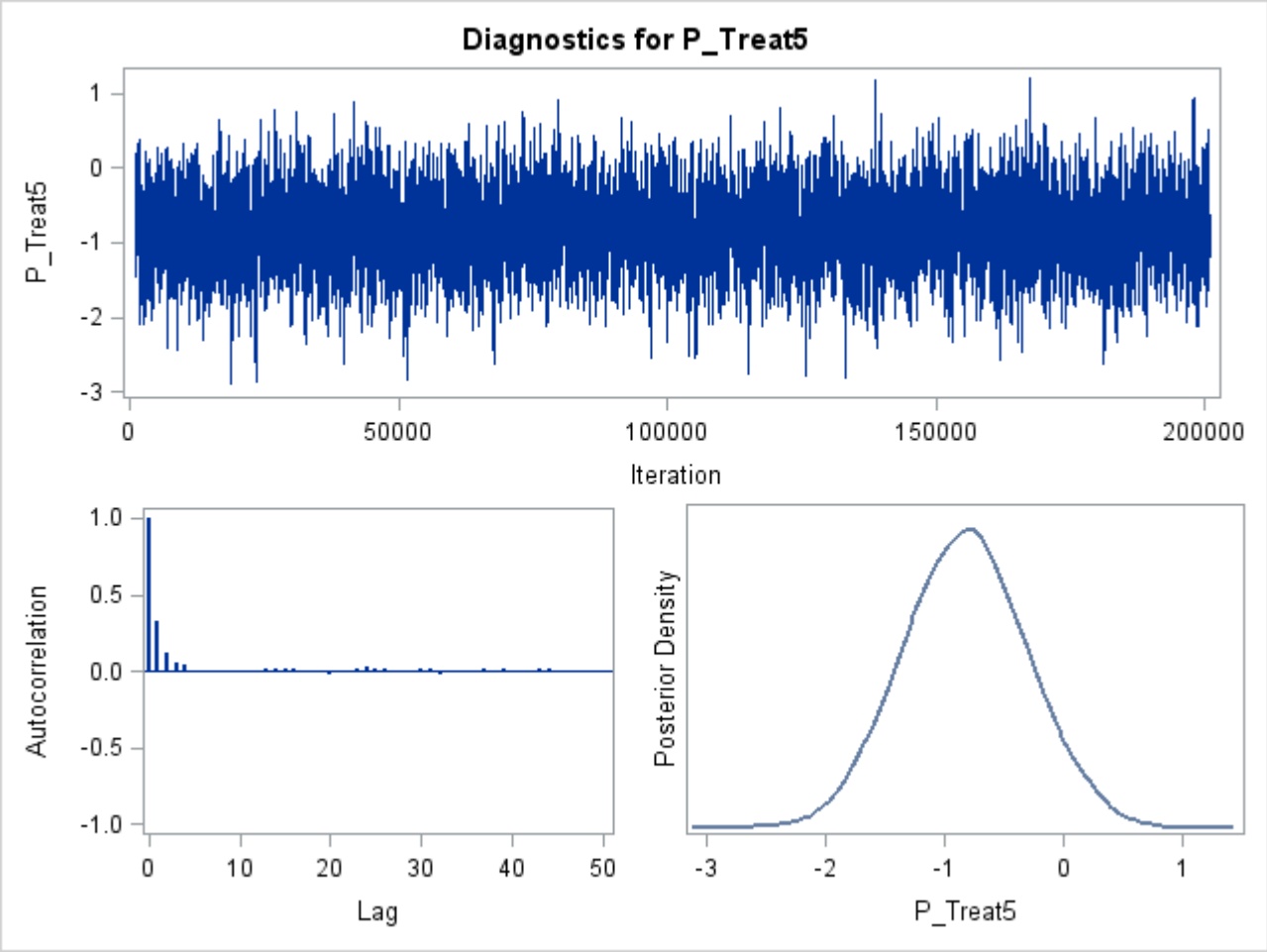
parms P_Study1-P_Study7 0 P_Treat2-P_Treat5 0;

Monte Carlo Standard Errors

Parameter	MCSE	Standard Deviation	MCSE/SD
P_Study1	0.00594	0.4299	0.0138
P_Study2	0.00373	0.2643	0.0141
P_Study3	0.00492	0.3602	0.0137
P_Study4	0.00750	0.5075	0.0148
P_Study5	0.00734	0.5284	0.0139
P_Study6	0.00705	0.5082	0.0139
P_Study7	0.00744	0.5133	0.0145
P_Treat2	0.00445	0.3321	0.0134
P_Treat3	0.00713	0.4872	0.0146
P_Treat4	0.00695	0.4824	0.0144
P_Treat5	0.00749	0.5240	0.0143

Bayesian often look to have MCSE/SD < 0.05.

Single PARMS and propcov=QUANEW



MCMC in SAS 9.3
Random statement makes code very easy!

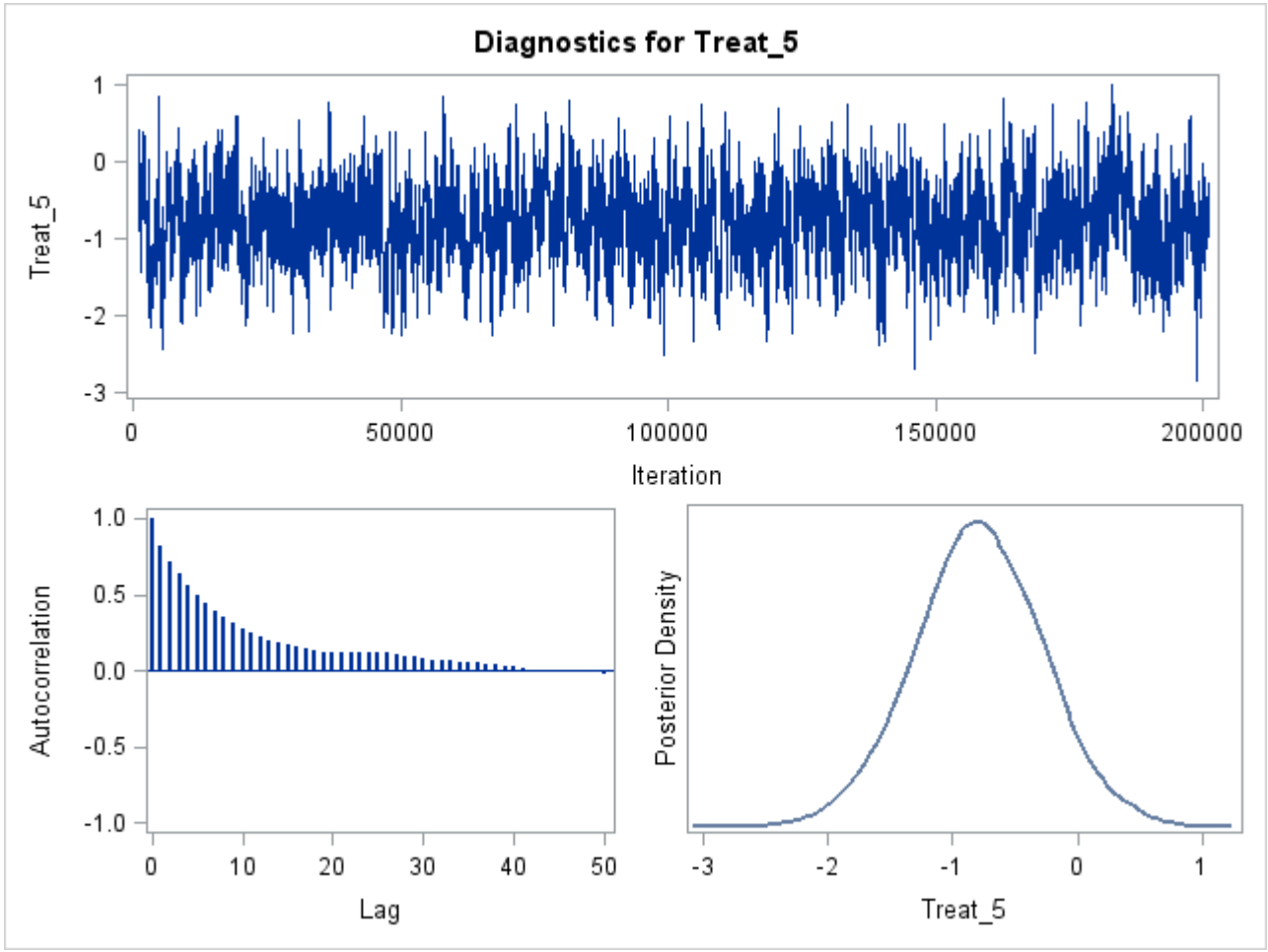
```
proc mcmc data=Parkinsons nmc=200000 thin=20
    seed=246810;

random Studyeffect ~general(0) subject=Study init=(0);

random Treat ~general(0) subject=Treatment  init=(0)
    zero=first  monitor=(Treat);

Mu=  Studyeffect + Treat ;

model Y ~ normal(mean=Mu, sd=SE);

run;
```

- A fixed effect is same as a random effect with fixed distribution (no parameters).
- Note. No need to know the number of levels.

Using RANDOM. MCSE not as good.

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Treat_3	10000	-0.4404	0.4788	-0.7605	-0.4360	-0.1178
Treat_2	10000	-1.8053	0.3304	-2.0297	-1.8030	-1.5815
Treat_4	10000	-0.4856	0.4731	-0.8029	-0.4833	-0.1696
Treat_5	10000	-0.7887	0.5151	-1.1272	-0.7894	-0.4423

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
Treat_3	0.0181	0.4788	0.0378
Treat_2	0.00419	0.3304	0.0127
Treat_4	0.0204	0.4731	0.0431
Treat_5	0.0216	0.5151	0.0419


Diagnostics for Treat_5
Treat_5
1
0
-1
-2
-3
0
50000
100000
150000
200000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
-3
-2
-1
0
1
Treat_5

Summary
[Normal data: Fixed effects – Bayesian]

- In SAS use the GENMOD procedure if you do not have any additional random effects.

- Theory says that with uninformative conjugate priors the results are identical to those from frequentist analysis.
  - Posterior means/medians match M.L estimates.
  - Credibility intervals match confidence intervals.

RANDOM EFFECTS MODEL

“Random effects” model

- Now we introduce the what is called the “random effects” model.
- Up until now the estimated overall treatment effect, estimates the average effect across this set of studies weighted by the size of each study.
  - If effect is same in all studies then this is a valid estimator.
  - If the effects (treatment differences) vary from study then this is still a valid estimator for this exact **weighting** of the individual differences in each trial.
- Here we will look at a more general *average across studies*.
- We introduce additional variability at the study level in terms of the average treatment effects.

The Statistical model

- Add random effect
- Study i and Arm k, with Treatment t(i,k)

$$Y_{ik} \sim N\left(\theta_{ik}, se_{ik}^2\right)$$

$$\theta_{ik} = \mu_i + \delta_{t(i,k)} + \eta_{ik}$$

where η_{ik} has zero mean, independent between studies with

$\text{Cov}(\eta_{ik} , \eta_{ih}) = \omega_{kh}$

[See Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC et al. Statistical approaches for conducting network meta-analysis in drug development, Pharmaceutical Statistics 2011, 10, 523-531]

The Statistical model

Usually cannot estimate the many parameters ω_{kh}.

- Symmetry is assumed leading to two possible options

where the i'th study has m_i arms (Ω is m_i by m_i).

1) $\omega_{kk} = \sigma^2/2$ and $\omega_{kh} = 0$ if $k \neq h$.

this is a simple diagonal matrix.

2) $\omega_{kk} = (m_i - 1)\,\sigma^2/2m_i = \sigma^2/2 - \sigma^2/2m_i$

and $\omega_{kh} = -\sigma^2/2m_i$ if $k \neq h$.

In this case Ω is not of full rank and

Var $(\eta_1 + \eta_2 + \ldots + \eta_{mi}) = 0$.

The Variance-covariance matrix Omega.

- In both cases we have

  $$\mathrm{Var}(\eta_k - \eta_h) = \sigma^2$$

  which is stable across studies however big.

- Model 1 is identical to Model 2 with additional simple random Study effect with variance $\sigma^2/2m$ for a study with m arms.
  - This is important when we decide whether solutions using (1) or (2) are equivalent or not.

The two forms.

- For 2 and 3 arm trials

(1)

$$\begin{bmatrix} \sigma^2/2 & 0 \\ 0 & \sigma^2/2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2/2 & 0 & 0 \\ 0 & \sigma^2/2 & 0 \\ 0 & 0 & \sigma^2/2 \end{bmatrix}$$

(2)

$$\begin{bmatrix} \sigma^2/4 & -\sigma^2/4 \\ -\sigma^2/4 & \sigma^2/4 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2/3 & -\sigma^2/6 & -\sigma^2/6 \\ -\sigma^2/6 & \sigma^2/3 & -\sigma^2/6 \\ -\sigma^2/6 & -\sigma^2/6 & \sigma^2/3 \end{bmatrix}$$

Means of the random effects within Study are aliased with the fixed effect for Study.

- Second version of Omega is obtained by simply subtracting the mean off the random effects within each trial.

$$\eta_i^* = \eta_i - \frac{1}{p}\sum_{j=1}^{p} \eta_j$$

- Option (1) has random effect on top of fixed effect.
- Option (2) obviates this complication for estimation.

- See Piepho H. P., Williams E. R., and Madden L. V.. 2012. *The Use of Two-Way Linear Mixed Models in Multitreatment Meta-Analysis.* Biometrics.

  for clear details of situations where these are equivalent.

Also contains useful references to early work, such as

De Hoog, F. R., Speed, T. P., and Williams, E. R. (1990). On a matrix identity associated with generalized least squares. *Linear Algebra and its Applications* ***127, 449–456.***

Easy way to use model (1)

- Use same approach as the Fixed effect model but add a random effect on every observation.

- Two possible ways ...
  - Fixed on RANDOM statement and estimated on REPEATED statement.
  - Vice-versa.

Specify known variances on RANDOM

```
proc mixed data= Parkinsons;
class Study Treatment Record;
model  Y = Study Treatment  /solution  ddfm=kr ;
random SE / subject=Study*Treatment;
parms 1 1 / HOLD=(1);
lsmeans Treatment / diff=control("1");
run;
```

Specify known variances on REPEATED

```
proc mixed data=Parkinsonstimes4 ;
class Study Treatment;
model Y= Study Treatment / solution ddfm=kr;
random intercept /subject=Study*Treatment ;
parms 1 1 / hold=(2);
weight Weight;
lsmeans Treatment / diff=control("1");
run;
```

There is no heterogeneity in these data.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SE	Study*Treatment	1.0000
Residual		0

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	Study*Treatment	0
Residual		1.0000

Easy way to use model (2)
... either frequentist or Bayesian.

- Within each Study set up three (maximum number of arms per study) random effects and then use weighted sums using weights
  - $(1-1/m) / \sqrt{2}$ on the diagonal
  - $(-1/m) / \sqrt{2}$ off the diagonal

  where m is number of arms for this study.

- $\text{Var} = \sigma^2 [(1-1/m)^2/2 + (m-1)/2m^2] = (m-1)\sigma^2/2m$
- $\text{Covariance} = \sigma^2 [(-2*(2(1-1/m) /m) + (m-2)2/m^2] = -2\sigma^2/m$

as required.

Weights ...

If m=2: $[1/2 , - 1/2, 0] /\sqrt{2}$ and $[- 1/2 , 1/2, 0]/\sqrt{2}$

If m=3: $[2/3, -1/3, -1/3]/\sqrt{2}$, $[-1/3, 2/3, -1/3]/\sqrt{2}$, etc.

Easy way to use model (2)
... either frequentist or Bayesian.

```
data Revised_data;
set Parkinsons;
by Study;
array x[3]x1-x3;
retain index;
drop i;
if  first.study  then index=0;
index=index+1;
do i=1 to 3;
     if i<= narm then x[i]=( (i=index) - (1/narm) ) / sqrt(2);
     else x[i]=0;
end;
run;
```

- (1-1/m) / √2 on the diagonal
- (-1/m) / √2 off the diagonal

Weights X1, X2 and X3

Record	Study	Treatment	Narm	x1	x2	x3	index
1	1	1	2	0.35	-0.35	0.00	1
2	1	3	2	-0.35	0.35	0.00	2
3	2	1	2	0.35	-0.35	0.00	1
4	2	2	2	-0.35	0.35	0.00	2
5	3	1	3	0.47	-0.24	-0.24	1
6	3	2	3	-0.24	0.47	-0.24	2
7	3	4	3	-0.24	-0.24	0.47	3
8	4	3	2	0.35	-0.35	0.00	1
9	4	4	2	-0.35	0.35	0.00	2
10	5	3	2	0.35	-0.35	0.00	1
11	5	4	2	-0.35	0.35	0.00	2
12	6	4	2	0.35	-0.35	0.00	1
13	6	5	2	-0.35	0.35	0.00	2
14	7	4	2	0.35	-0.35	0.00	1
15	7	5	2	-0.35	0.35	0.00	2

Easy way to use model (2)
... either frequentist or Bayesian.

```
proc mixed data=Revised_Data;
class Study Treatment;
Model Y =Study Treatment / ddfm=kr;
random X1 X2 X3/ subject=study type=toep(1);
weight Weight;
parms 1 1 /hold=(2);
lsmeans Treatment / diff=control("1“);
 run;
```

$$G = \begin{matrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{matrix}$$

Not TYPE=VC would have separate variances and be wrong.

ML estimate of random effect variance is zero.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	Study	0
Residual		1.0000

REML

- By conditioning on the estimators of the fixed effect parameters, the REML likelihoods are the same for (1) and (2).
  - So estimates and their SEs are the same.

- Similarly, Bayesian analysis with flat priors for Study fixed effects give identical posteriors for the two different Omega models.

Using proc MIXED on full data set Model (1) using REML.

```
proc mixed data=Full;
class Study Treatment;
model Response=Study Treatment /ddfm=kr;
lsmeans Treatment / diff=control("1");
random Treatment / subject=Study;
repeated /subject=FullRec group=Study*Treatment;
run;
```

- [e.g. Whitehead §5.8.2]

```
random Treatment * Study;
random Treatment / subject=Study;
random Intercept  / subject=Treatment*Study;
```
← All three are equivalent

There is no extra variability at study level.

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Study*Treatment			0
Residual	FullRec	Study*Treatment 1 1	13.6307
Residual	FullRec	Study*Treatment 1 3	18.2842
Residual	FullRec	Study*Treatment 2 1	13.6831
Residual	FullRec	Study*Treatment 2 2	11.5551
Residual	FullRec	Study*Treatment 3 1	19.3197
Residual	FullRec	Study*Treatment 3 2	18.4155
Residual	FullRec	Study*Treatment 3 4	18.4420
Residual	FullRec	Study*Treatment 4 3	9.0001
Residual	FullRec	Study*Treatment 4 4	9.0005
Residual	FullRec	Study*Treatment 5 3	9.0203
Residual	FullRec	Study*Treatment 5 4	9.0621
Residual	FullRec	Study*Treatment 6 4	5.3275
Residual	FullRec	Study*Treatment 6 5	4.7449
Residual	FullRec	Study*Treatment 7 4	6.1413
Residual	FullRec	Study*Treatment 7 5	8.9179

No observable variation at Study level

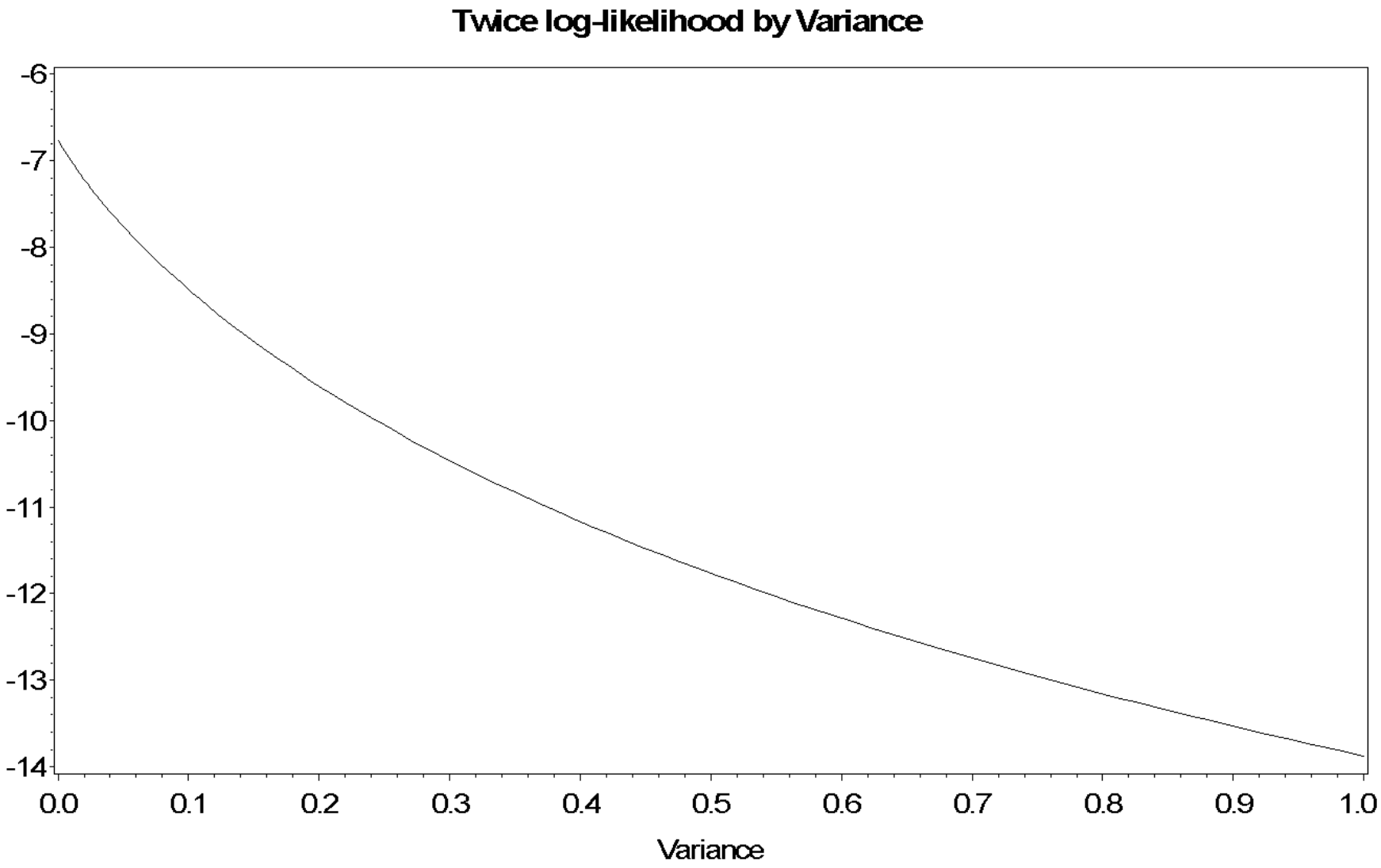
- Frequentist accepts this and effectively opts for the Fixed effects analysis.
- The Bayesian believes his prior and carries on.
- Note there were only 4 d.f. to estimate this study level variation.

Lots of fixed effect parameters

... but little data.

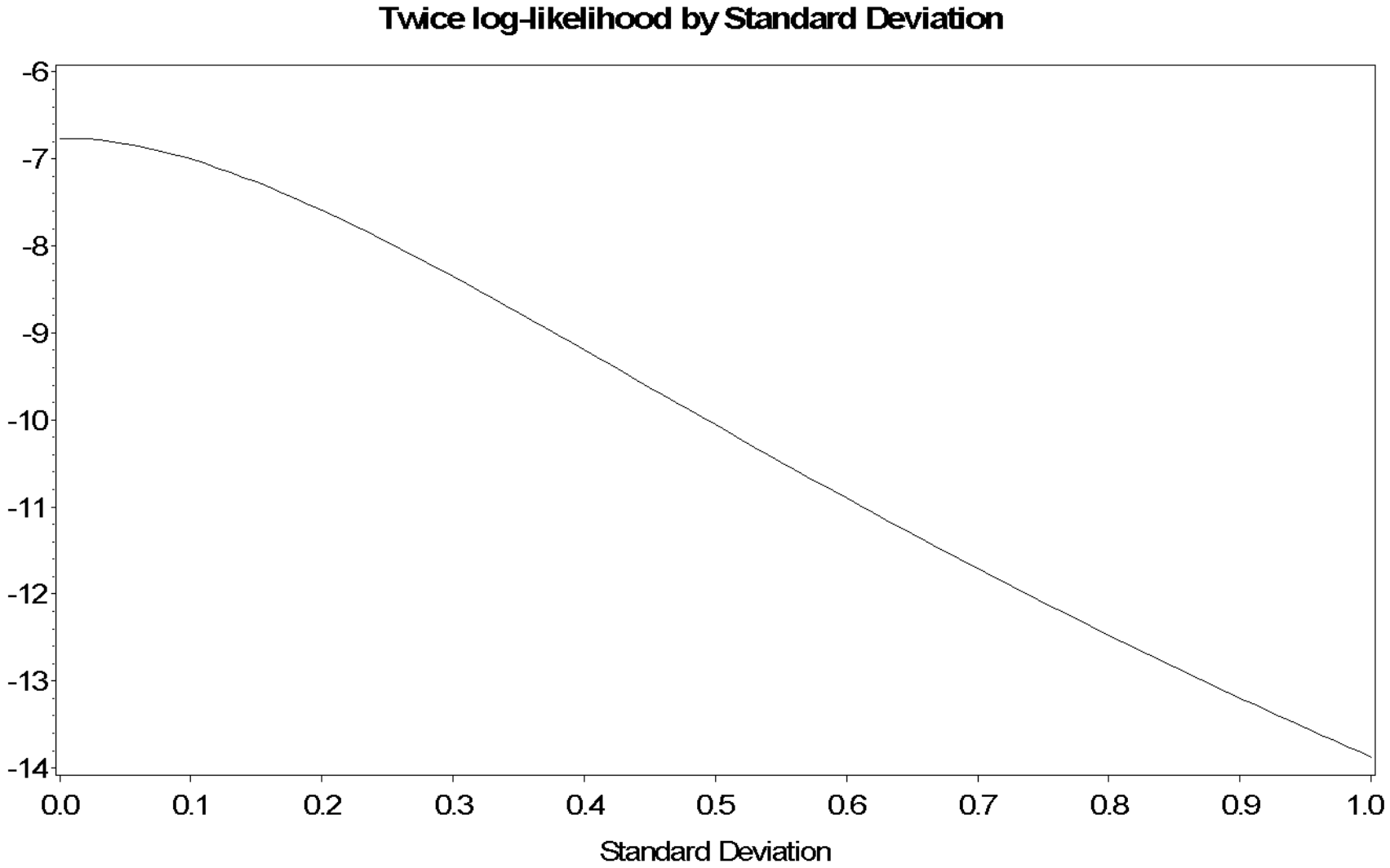
Study/Treatment	1	2	3	4	5	Margin
1	-1.22		-1.53			μ_1
2	-0.7	-2.4				μ_2
3	-0.3	-2.6		-1.2		μ_3
4			-0.24	-0.59		μ_4
5			-0.73	-0.18		μ_5
6				-2.2	-2.5	μ_6
7				-1.8	-2.1	μ_7
Margin	$\delta_1=0$	δ_2	δ_3	δ_4	δ_5	

Profile likelihood for Variance.



Code for this plot supplied in course materials.

Profile likelihood for SD.



Code for this plot supplied in course materials.

Version (1) of Omega matrix.

- Same model as for fixed effect except that

$$\mathrm{Var}(Y_{ik}) = se_{ik}^2 + \sigma^2/2$$

where

se_{ik}^2 is known and σ^2 needs to be estimated.

Priors for SD or Variance

- For a Bayesian model we need prior for σ^2

Uniform for SD is commonly used for variance components which are not at lowest stratum level.

$$d\sigma = \sigma\, d(\log \sigma) = d(\sigma^2)/2\sigma$$

MCMC procedure:

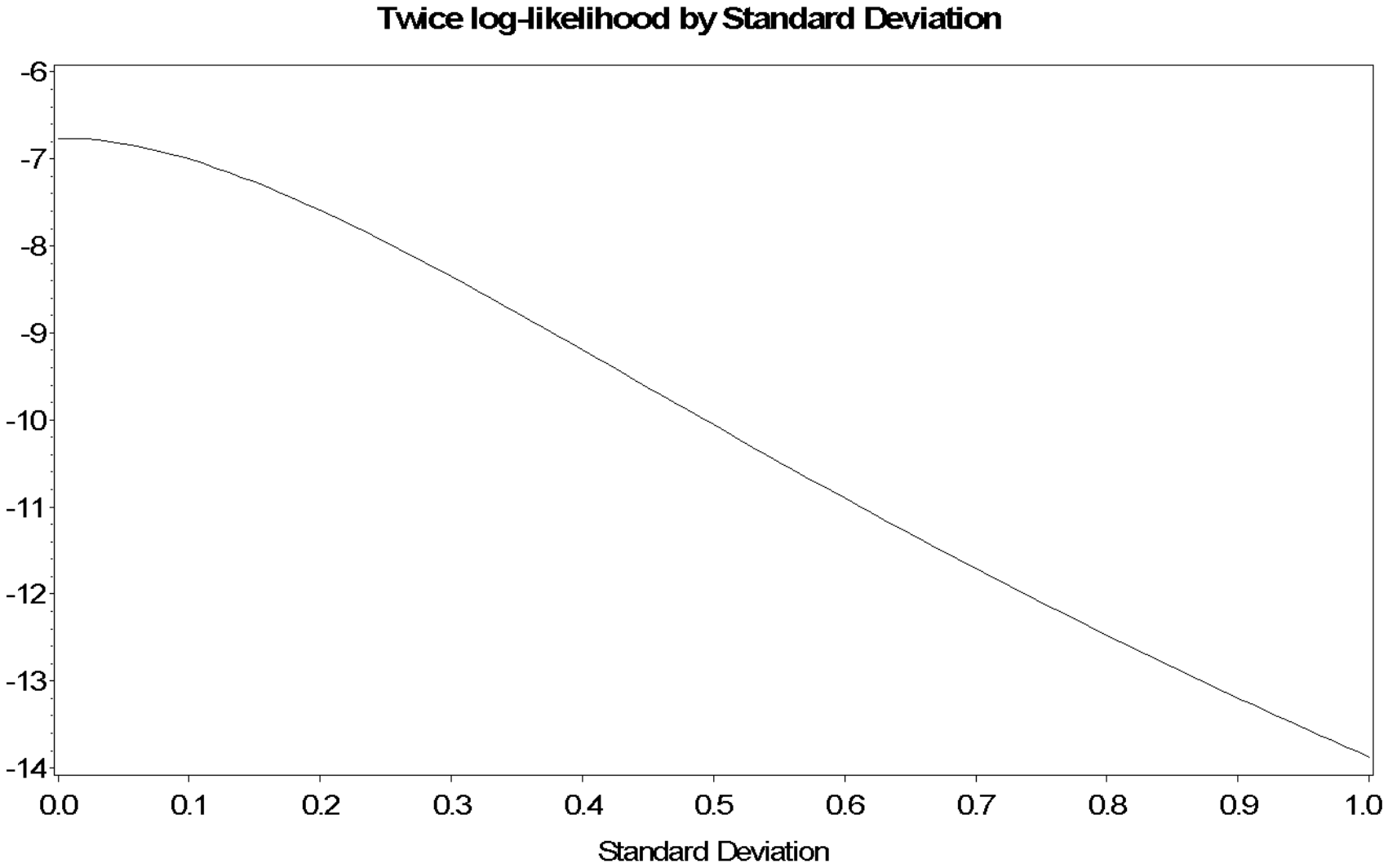
Prior sd ~ general(0) OR

prior sd ~ uniform(0.001,5)

Winbugs:

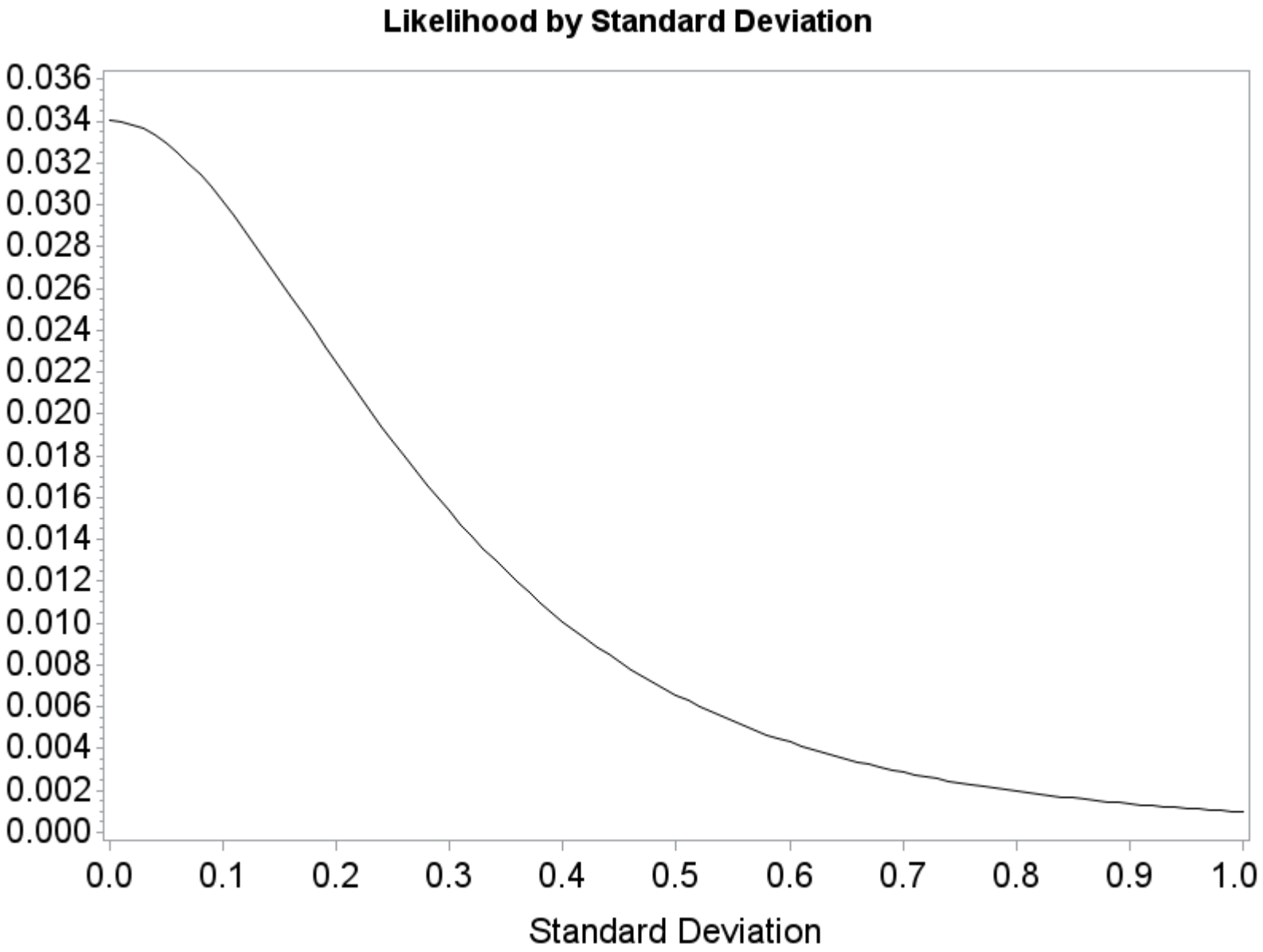
sd ~ dunif(0,5)

Note how flat prior on SD will average over range 0 up to about 0.3.



Code for this plot supplied in course materials.

With flat SD prior, this is what we integrate over.



Winbugs

- Most available code uses version (2) of Ω matrix.

- But this is not necessary with flat priors for the study effects and linear link function.

Winbugs code from Nice for random effects.

Sofia Dias, Nicky J Welton, Alex J Sutton, AE Ades.

```
# Normal likelihood, identity link.   Random effects model for multi-arm trials
model{ # *** PROGRAM STARTS
for(i in 1:ns){ # LOOP THROUGH STUDIES
    w[i,1] <- 0 # adjustment for multi-arm trials is zero for control arm
    delta[i,1] <- 0 # treatment effect is zero for control arm
    mu[i] ~ dnorm(0,.0001) # vague priors for all trial baselines
    for (k in 1:na[i]) { # LOOP THROUGH ARMS
        var[i,k] <- pow(se[i,k],2) # calculate variances
        prec[i,k] <- 1/var[i,k] # set precisions
        y[i,k] ~ dnorm(theta[i,k],prec[i,k]) # normal likelihood
        theta[i,k] <- mu[i] + delta[i,k] # model for linear predictor
        dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k] #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na[i]]) # summed residual deviance contribution for this trial
    for (k in 2:na[i]) {                    ← This is over complicated!
        delta[i,k] ~ dnorm(md[i,k],taud[i,k]) # trial-specific LOR distributions
        md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k] # mean of treat effects distributions (with multi-arm trial correction)
        taud[i,k] <- tau *2*(k-1)/k # precision of treat effects distributions (with multi-arm trial correction)
        w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]]) # adjustment for multi-arm RCTs
        sw[i,k] <- sum(w[i,1:k-1])/(k-1) # cumulative adjustment for multi-arm trials
    }       }
totresdev <- sum(resdev[]) #Total Residual Deviance
d[1]<-0 # treatment effect is zero for reference treatment
for (k in 2:nt){ d[k] ~ dnorm(0,.0001) } # vague priors for treatment effects
sd ~ dunif(0,5) # vague prior for between-trial SD.
tau <- pow(sd,-2) # between-trial precision = (1/between-trial variance)
}
```

Much simpler code, subtracting mean random effect (effectively Omega (2)).

```
# Normal likelihood, identity link
# Random effects model for multi-arm trials
model{ #
for(i in 1:ns){
  mu[i] ~ dnorm(0,.0001) # vague priors for all trial baselines
  for (k in 1:na[i]) {
    var[i,k] <- pow(se[i,k],2) # calculate variances
    prec[i,k] <- 1/var[i,k] # set precisions
    y[i,k] ~ dnorm(theta[i,k],prec[i,k]) # normal likelihood
    theta[i,k] <- mu[i] + d[t[i,k]]  +delta[i,k] - dsum[i] # model for linear predictor
        delta[i,k] ~ dnorm(0,tau)
    dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k] #Deviance contribution
  }
        dsum[i] <- sum(delta[i,1:na[i]])
        resdev[i] <- sum(dev[i,1:na[i]]) # summed residual deviance
}
totresdev <- sum(resdev[]) #Total Residual Deviance
d[1]<-0 # treatment effect is zero for reference treatment
for (k in 2:nt){ d[k] ~ dnorm(0,.0001) } # vague priors for treatment effects
  sd ~ dunif(0,5) # vague prior for between-trial SD.
  tau <- pow(sd,-2) # between-trial precision = (1/between-trial variance)
}
```

The NICE results (Bayesian)

Table A9 Parkinson example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Treatments 2 to 5 (d_{12} to d_{15}) relative to Placebo, absolute effects of Placebo (T_1) and treatments 2 to 5 (T_2 to T_5), heterogeneity parameter τ and model fit statistics for different data types.

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
	Arm-level data: Example 5							
d_{12}	-1.81	0.33	-1.81	(-2.46,-1.16)	-1.85	0.54	-1.84	(-2.91,-0.85)
d_{13}	-0.47	0.49	-0.47	(-1.43,0.49)	-0.50	0.66	-0.50	(-1.78,0.75)
d_{14}	-0.52	0.48	-0.52	(-1.46,0.43)	-0.53	0.65	-0.53	(-1.77,0.71)
d_{15}	-0.82	0.52	-0.82	(-1.84,0.22)	-0.83	0.80	-0.83	(-2.35,0.69)
T_1	-0.73	0.22	-0.73	(-1.16,-0.30)	-0.73	0.22	-0.73	(-1.16,-0.30)
T_2	-2.54	0.40	-2.54	(-3.32,-1.76)	-2.58	0.58	-2.57	(-3.72,-1.50)
T_3	-1.21	0.53	-1.20	(-2.25,-0.15)	-1.23	0.70	-1.23	(-2.57,0.10)
T_4	-1.25	0.53	-1.25	(-2.28,-0.21)	-1.26	0.69	-1.26	(-2.57,0.05)
T_5	-1.55	0.57	-1.55	(-2.66,-0.43)	-1.57	0.83	-1.56	(-3.14,0.02)
τ	-	-	-	-	0.40	0.43	0.28	(0.01,1.55)

Winbugs using Omega (1)
... and masses of iterations.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
d[2]	-1.848	0.531	0.003018	-2.909	-1.838	-0.8604	10000	990001
d[3]	-0.4978	0.6629	0.006444	-1.777	-0.4927	0.7522	10000	990001
d[4]	-0.5311	0.6489	0.007535	-1.781	-0.5287	0.7055	10000	990001
d[5]	-0.8312	0.8035	0.009799	-2.364	-0.8293	0.6862	10000	990001
sd	0.2822	0.3166	0.004325	0.008667	0.1925	1.095	10000	990001

[Note sd is $\sqrt{2}$ smaller due to parameterisation.]

theta[i,k] <- mu[i] + d[t[i,k]] +delta[i,k]

delta[i,k] ~ dnorm(0,tau)

MCMC procedure in SAS

- Tends to get stuck if we do not keep SD away from zero.

```
prior sd ~ uniform(0.001,5) ;
```

OR better work on the logsd scale

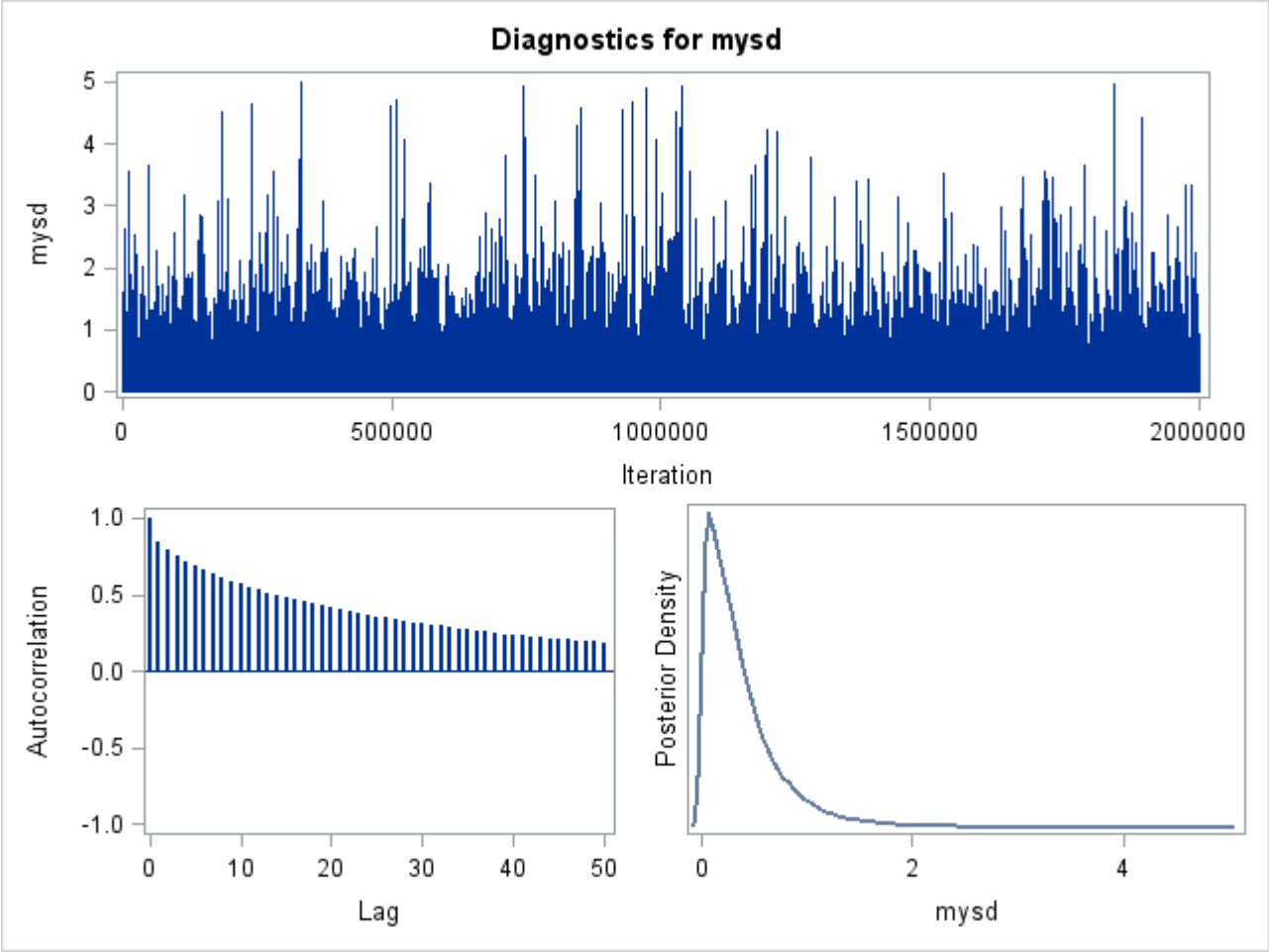
```
prior logsd ~ general(logsd, upper=log(5)) ;
```

Omega (1): Easy way is add on variance.

```
ods rtf file="&studydir.\output\ex1";
ods graphics on;
proc mcmc data=ParkBayes propcov=quanew nmc=2000000  thin=10 monitor=(P_Treat logsd mysd
      ) seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
parms P_Study1-P_Study7 0  P_Treat2-P_Treat5 0 ;
prior P_Study1-P_Study7  P_Treat2-P_Treat5 ~ general(0);
parms logsd 0;
prior logsd ~ general(logsd, upper=log(5));
mysd=exp(logsd);
P_Treat[1]=0;
Mu=  P_Study[Study] + P_Treat[Treatment] ;
v=Var+mysd*mysd/2;
model Y ~ normal(mean=Mu, var=v);
run;
```

Same as Nice results

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
P_Treat1	200000	0	0	0	0	0
P_Treat2	200000	-1.8505	0.5436	-2.1345	-1.8399	-1.5498
P_Treat3	200000	-0.4990	0.6781	-0.8872	-0.4928	-0.1055
P_Treat4	200000	-0.5280	0.6487	-0.9103	-0.5248	-0.1427
P_Treat5	200000	-0.8302	0.8007	-1.2696	-0.8255	-0.3831
logsd	200000	-1.4638	1.2239	-2.0871	-1.2810	-0.6448
mysd	200000	0.4022	0.4350	0.1240	0.2778	0.5248


Diagnostics for mysd
mysd
5
4
3
2
1
0
0
500000
1000000
1500000
2000000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
2
4
mysd

- But this does not generalise to case of Binary data.

- Need to introduce the random effect directly.
  - That will generalise.

Omega (1) and random effect. SAS 9.2

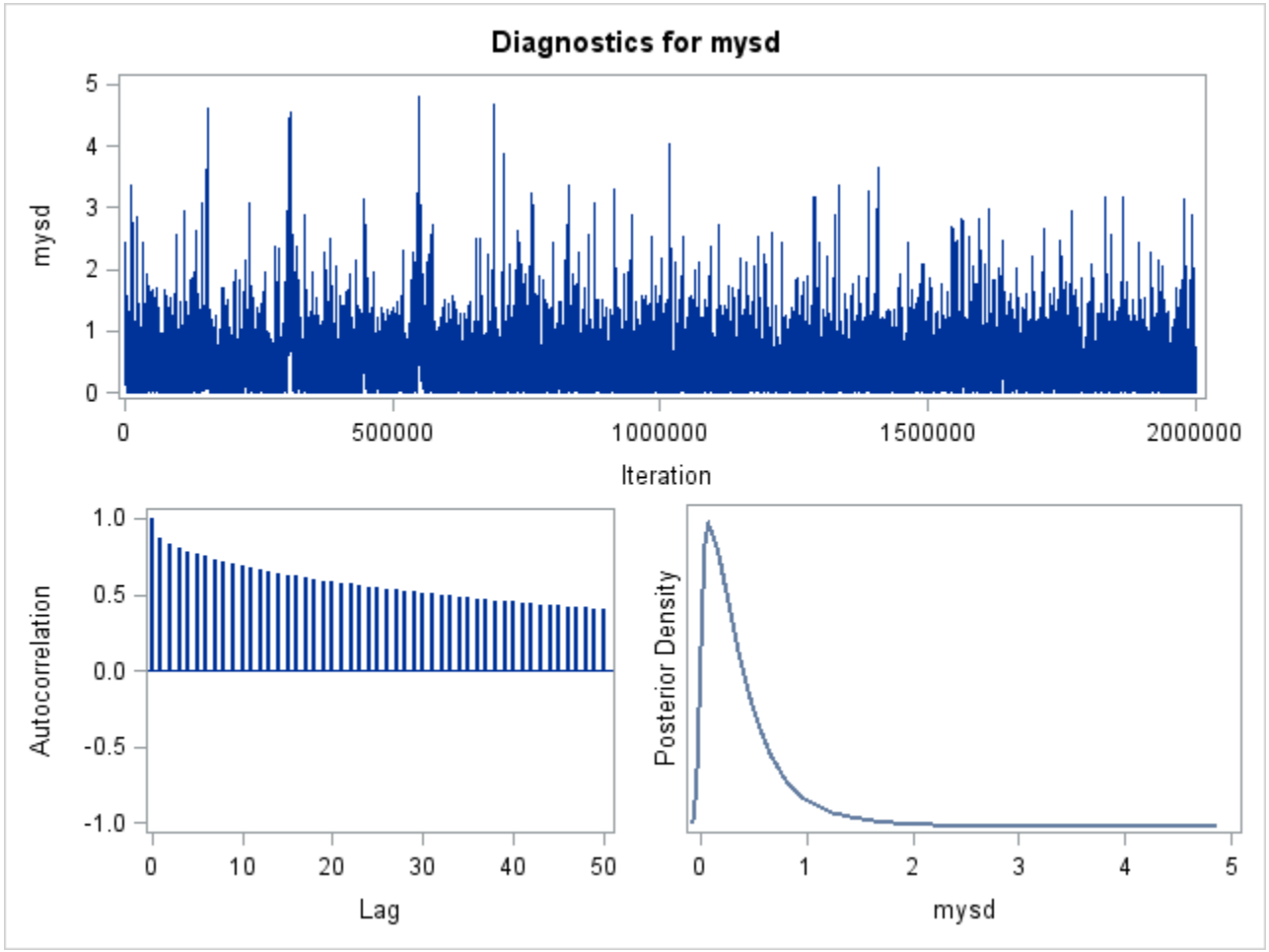
```
proc mcmc data=Parkinsons  nmc=2000000 thin=10 monitor=(P_Treat logsd mysd ) seed=246810;
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
array RE[15];
parms P_Study1-P_Study7 0  P_Treat2-P_Treat5 0 ;
prior P_Study1-P_Study7  P_Treat2-P_Treat5 ~ general(0);
parms logsd 0;
prior logsd ~ general(logsd,upper=log(5));
mysd=exp(logsd)/sqrt(2);
P_Treat[1]=0;

parms RE: /slice;
prior RE: ~ normal(0,sd=mysd);
Mu=  P_Study[Study] + P_Treat[Treatment] + RE[record];
model Y ~ normal(mean=Mu, SD=SE);
run;
```

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
P_Treat1	200000	0	0	0	0	0
P_Treat2	200000	-1.8746	0.5162	-2.1502	-1.8501	-1.5638
P_Treat3	200000	-0.5412	0.6339	-0.9129	-0.5214	-0.1367
P_Treat4	200000	-0.5871	0.6087	-0.9552	-0.5692	-0.1929
P_Treat5	200000	-0.9059	0.7447	-1.3229	-0.8787	-0.4505
mysd	200000	0.3867	0.3833	0.1248	0.2772	0.5212

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
P_Treat1	0	0	.
P_Treat2	0.0107	0.5162	0.0207
P_Treat3	0.0166	0.6339	0.0262
P_Treat4	0.0179	0.6087	0.0293
P_Treat5	0.0239	0.7447	0.0321
mysd	0.0116	0.3833	0.0303

/SLICE option on PARMS statement.

- The /SLICE uses a separate sweep for each random effect.
- Uses a slice sampler to sample.
- Both the use of SLICE and the increased number of sweeps of the data make this approach take very much longer.
  - 15 minutes, while the following SAS 9.3 code takes 6 seconds.
- But it does help the mixing of the SD parameter.


Diagnostics for mysd
mysd
5
4
3
2
1
0
0
500000
1000000
1500000
2000000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
1
2
3
4
5
mysd

But even easier in SAS 9.3

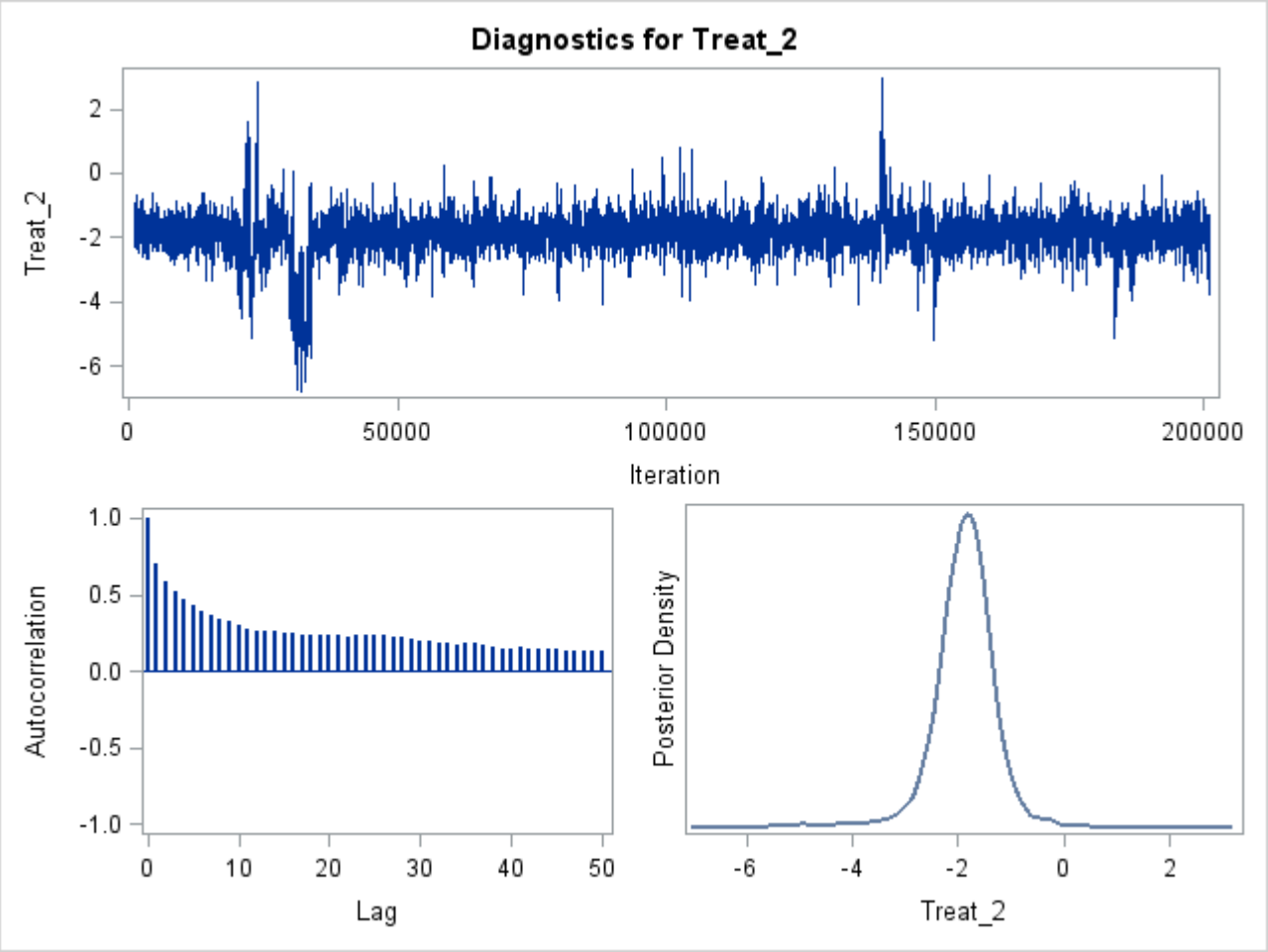
```
proc mcmc data=Parkinsons  nmc=200000 nthin=20 seed=246810
    monitor=(mysd);

random Studyeffect ~general(0) subject=Study init=(0) ;

random Treat ~general(0) subject=Treatment  init=(0) zero=first
    monitor=(Treat);

parms logsd 0;

prior logsd ~ general(logsd, upper=log(5));

mysd=exp(logsd);

random RE ~normal(0,sd=mysd/sqrt(2)) subject=_OBS_ init=(0);

Mu=  Studyeffect + Treat +RE;

model Y ~ normal(mean=Mu, sd=SE);

run;
```

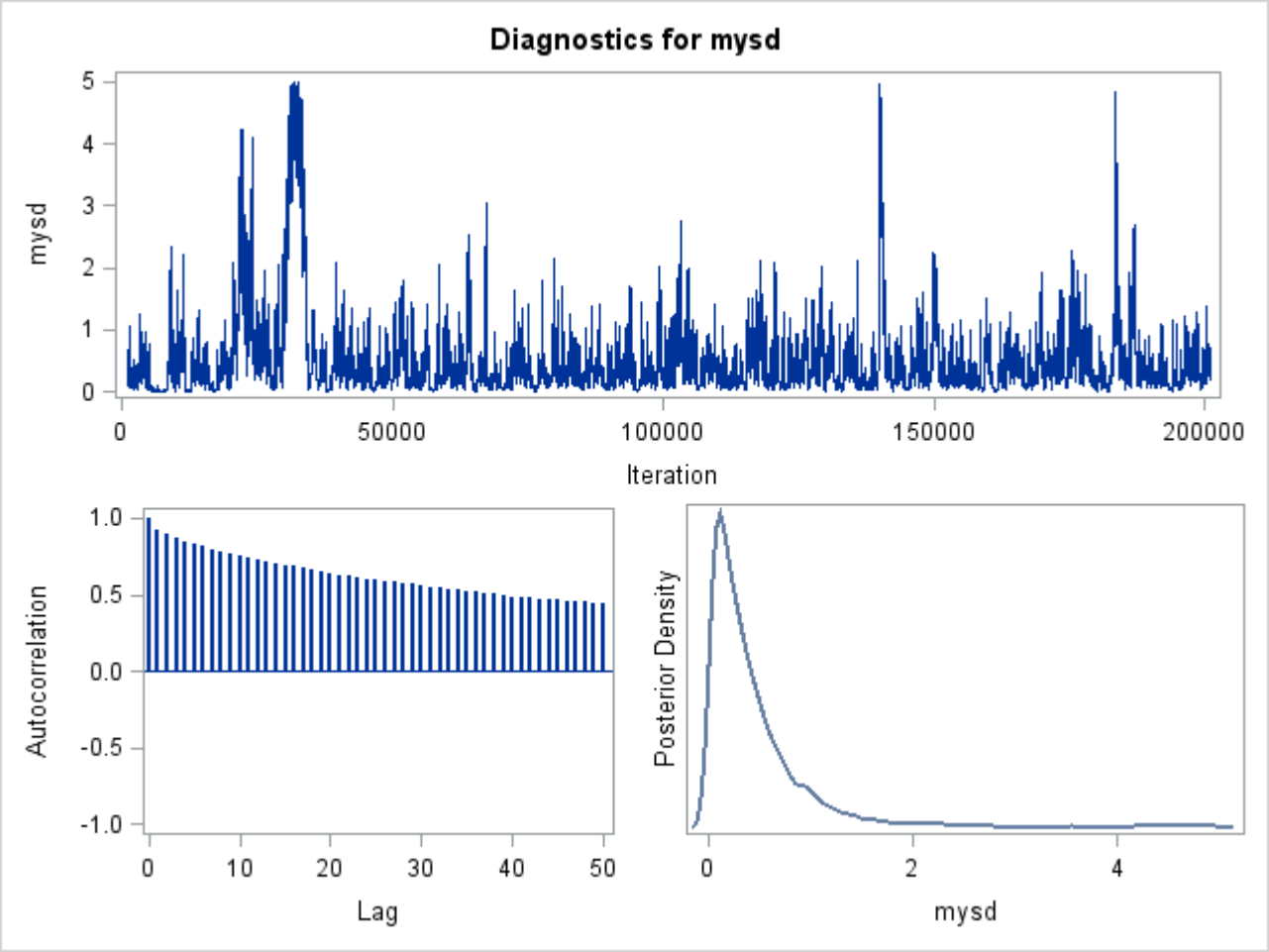
Results. Fast but need more iterations.

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
mysd	10000	0.4940	0.6525	0.1362	0.3019	0.5878
Treat_3	10000	-0.5009	0.7351	-0.9027	-0.4901	-0.0841
Treat_2	10000	-1.8846	0.6800	-2.1703	-1.8466	-1.5430
Treat_4	10000	-0.6130	0.8344	-0.9552	-0.5455	-0.1464
Treat_5	10000	-0.9796	1.1094	-1.3342	-0.8442	-0.3859

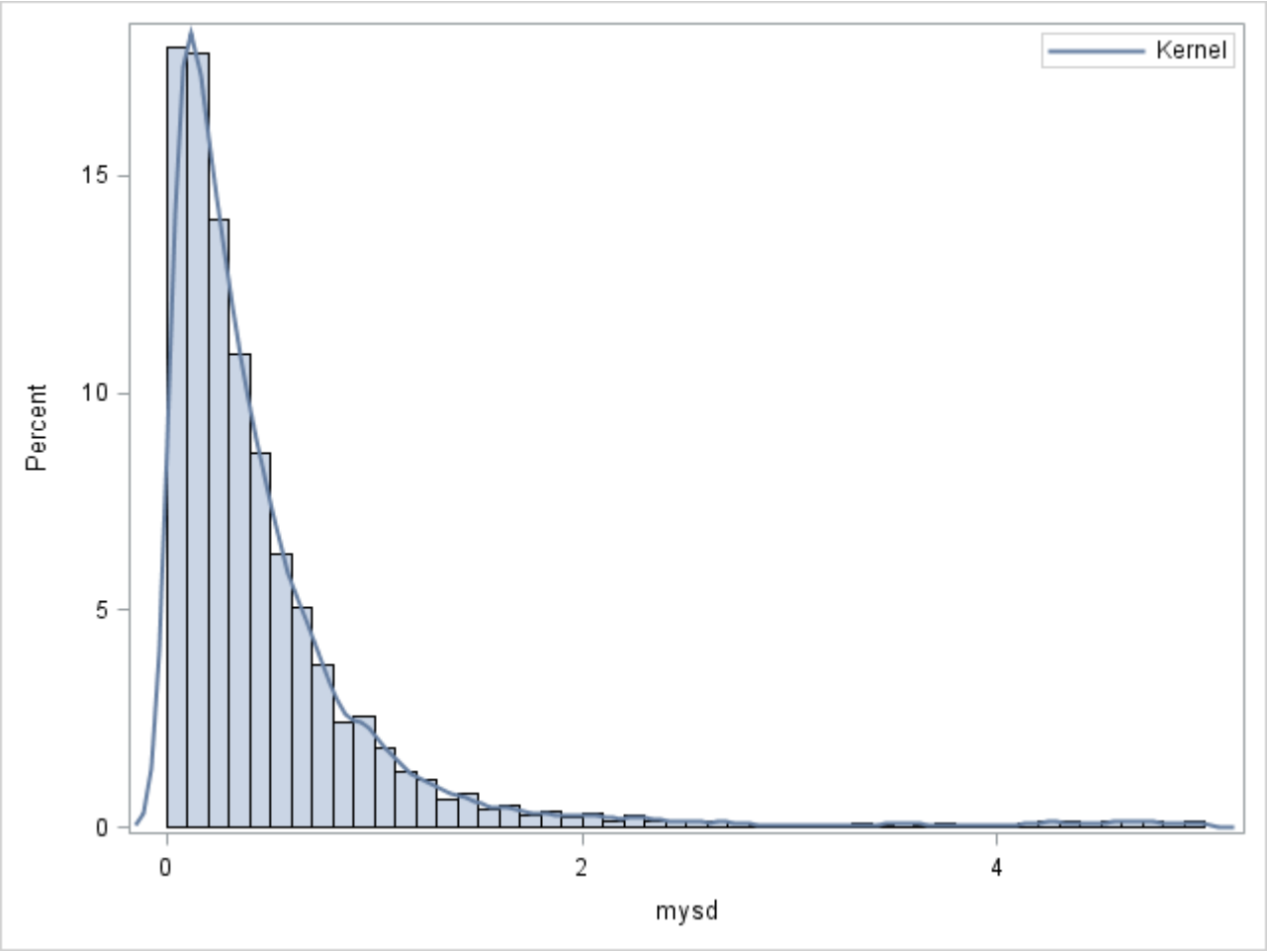
Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
mysd	0.0697	0.6525	0.1068
Treat_3	0.0380	0.7351	0.0517
Treat_2	0.0462	0.6800	0.0679
Treat_4	0.0882	0.8344	0.1057
Treat_5	0.1276	1.1094	0.1150

Treatment not mixing as well




Diagnostics for mysd
mysd
5
4
3
2
1
0
0
50000
100000
150000
200000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
2
4
mysd

Underlying histogram.



Use SGPLOT and not K3D

```
proc mcmc data=Parkinsons  nmc=200000 nthin=20 seed=246810
     monitor=(mysd) outpost=outp1;

... ...

run;


proc sgplot data=outp1;
  histogram mysd /binstart=0.05 binwidth=0.1;
  density mysd / type=kernel;
  keylegend / location=inside position=topright;
run;
```

- Need binstart= and binwidth= to get cell to start at zero.

- May need to include a lower limit for SD as well.
  - Monitor the diagnostic graphs.

```
parms logsd 0;
prior logsd ~ general(logsd,
        lower=log(0.001), upper=log(5));
```

Omega (2)

- ... or 15 random effects and use our X1, X2, X3 trick.

- Here the propcov=quanew trick works poorly.

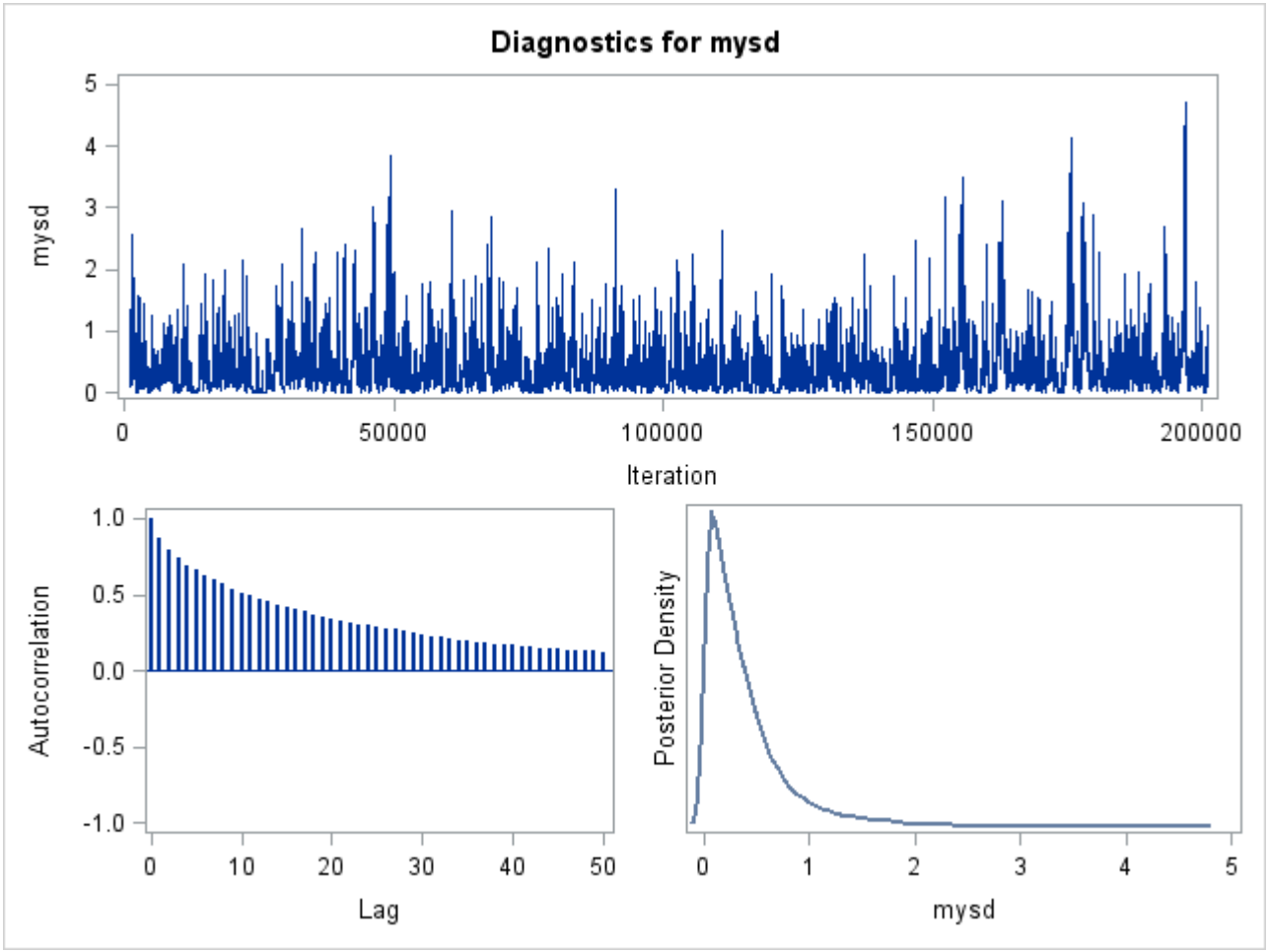
- For random effect we use
  - SLICE option on the PARMS statement for the random effects in SAS 9.2.
  - RANDOM statement in SAS 9.3

Set up the three variables X1, X2 and X3.

```
data Parkbayes;
set Parkinsons;
by study;
retain offset 0 lastnarm 0;
array x[3]x1-x3;
retain index;
if  first.study  then index=0;
index=index+1;
drop i lastnarm;
if first.study then offset=offset+lastnarm;
lastnarm=narm;
do i=1 to 3;
     if i<= narm then x[i]=((i=index)-(1/narm)) / sqrt(0.5);
     else x[i]=0;
end; run;
```

```
proc mcmc data=ParkBayes  ntu=1000 nmc=200000 nthin=10 seed=246810 monitor=(P_Treat2-
      P_Treat5 logsd mysd);
array P_Study[7] P_Study1-P_Study7;
array P_Treat[5] P_Treat1-P_Treat5;
array P_rand[15] P_Rand1-P_Rand15;
array x[3] x1-x3;
parms P_Study1-P_Study7 0  P_Treat2-P_Treat5 0 ;
parms P_rand1-P_rand15 /slice;
parms logsd 0;
prior logsd ~ general(logsd, upper=log(5));
mysd=exp(logsd);
prior P_Rand:  ~ normal(0,sd=mysd);
prior P_Study1-P_Study7 ~ general(0);
prior P_Treat2-P_Treat5 ~ general(0);
P_Treat[1]=0;
sum=0;
do i=1 to narm;
      sum=sum+x[i]*P_Rand[offset+i];
end;
Mu=  P_Study[Study] + P_Treat[Treatment] + sum;
model Y ~ normal(mean=Mu, sd=SE); run;
```

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
P_Treat2	20000	-1.8489	0.5341	-2.1292	-1.8413	-1.5492
P_Treat3	20000	-0.5111	0.6358	-0.8839	-0.4975	-0.1217
P_Treat4	20000	-0.5565	0.6334	-0.9201	-0.5411	-0.1649
P_Treat5	20000	-0.8819	0.7827	-1.2817	-0.8480	-0.4211
logsd	20000	-1.4984	1.1884	-2.1647	-1.3213	-0.6826
mysd	20000	0.3880	0.4171	0.1148	0.2668	0.5053


Diagnostics for mysd
mysd
5
4
3
2
1
0
0
50000
100000
150000
200000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
1
2
3
4
5
mysd

Using RANDOM statement
Needs lower bound for logsd prior.

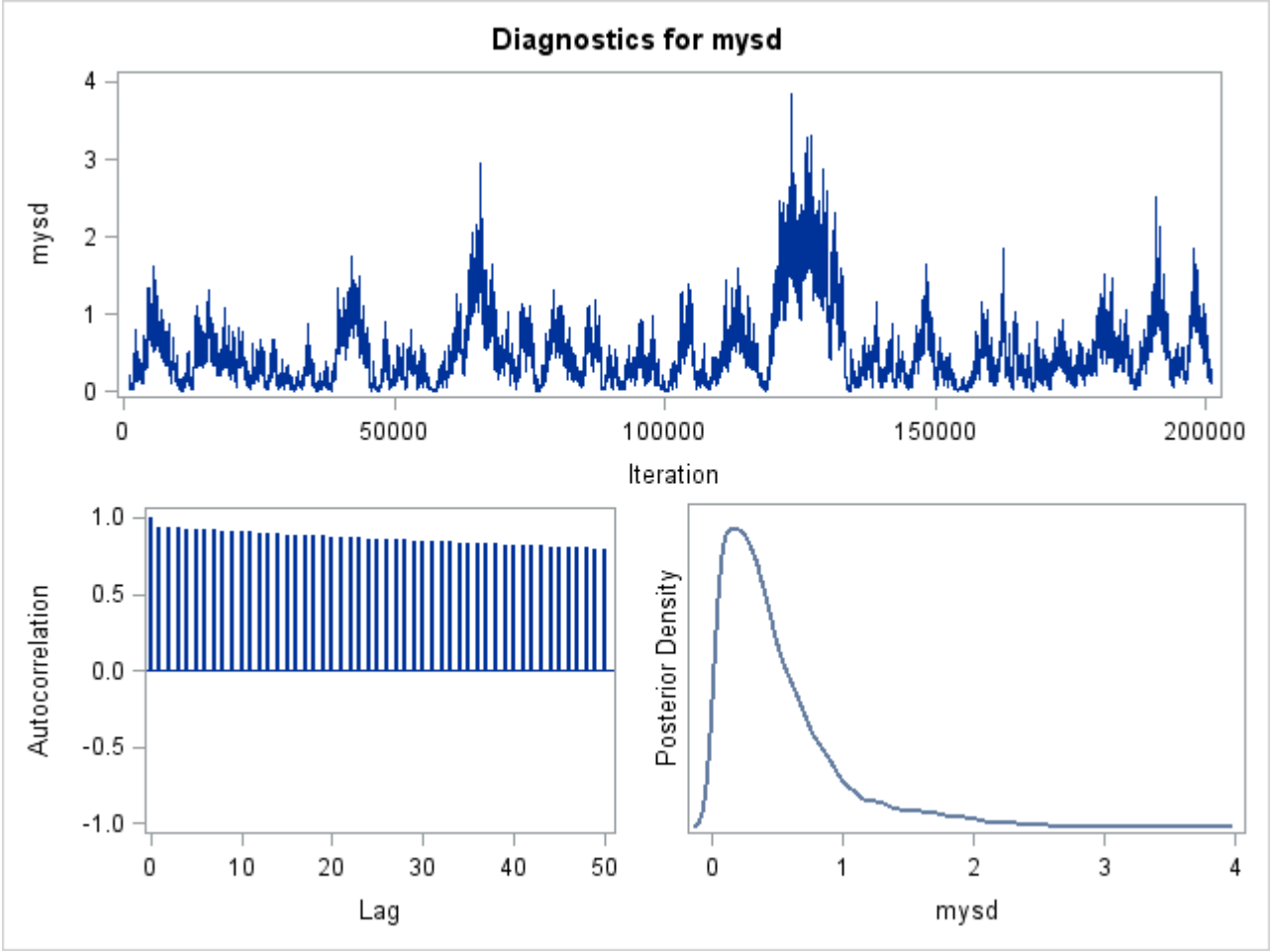
```
proc mcmc data=ParkBayes  nmc=200000 thin=10 seed=246810 monitor=(logsd mysd);
array x[3] x1-x3;
array P_Rand[3];
array zero[3] (0,0,0);
parms logsd 0;
prior logsd ~ general(logsd,lower=log(0.01) upper=log(5));
mysd=exp(logsd);

random Studyeffect ~general(0) subject=Study init=(0) ;
random Treat ~general(0) subject=Treatment  init=(0) zero=first monitor=(Treat);
random P_Rand ~ mvnar(mean=zero, sd=mysd, 0) subject=study ;
sum=0;
do i=1 to narm;
      sum=sum+x[i]*P_Rand[i];
end;
Mu=  Studyeffect + Treat + sum;
model Y ~ normal(mean=Mu, sd=SE); run;
```

Random effect SD is largish

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
logsd	20000	-1.1461	1.0121	-1.7023	-1.0082	-0.4418
mysd	20000	0.4825	0.4326	0.1823	0.3649	0.6429
Treat_3	20000	-0.5309	0.7811	-0.8958	-0.4986	-0.0821
Treat_2	20000	-1.8723	0.5355	-2.1673	-1.8537	-1.5489
Treat_4	20000	-0.5416	0.7279	-0.9362	-0.5306	-0.0948
Treat_5	20000	-0.8274	0.9531	-1.3041	-0.8392	-0.3376

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
logsd	0.1177	1.0121	0.1163
mysd	0.0669	0.4326	0.1547
Treat_3	0.0890	0.7811	0.1140
Treat_2	0.0345	0.5355	0.0644
Treat_4	0.0757	0.7279	0.1040
Treat_5	0.1093	0.9531	0.1147

Logsd and SD are not mixing well



Summary for Bayesian
[Normal data: Random effects]

- Use code

  parms logsd 0;

  prior logsd ~ general(logsd,lower=log(0.01) upper=log(5));

  mysd=exp(logsd);

  but lower limit may not be needed, especially when heterogeneity exists.

- Random effect introduces lots of additional parameters, one for each record in data set.
  - For Normal data simply add variance onto the fixed known residual and use same code as for fixed effects.
  - We will need to include random effect specifically when we move to non-Normal data.

Summary
[Normal data: Random effects]

- Different results between Frequentist and Bayesian, especially if small observed variability at study level.
  - This is to be expected.
  - Main difference is the increase in width of Credibility interval compared to confidence interval.

- For Linear link both models (1) and (2) for Omega give same results.
  - When using REML (Frequentist).
  - When using flat priors for Study effect (Bayesian).

Variability at top stratum.

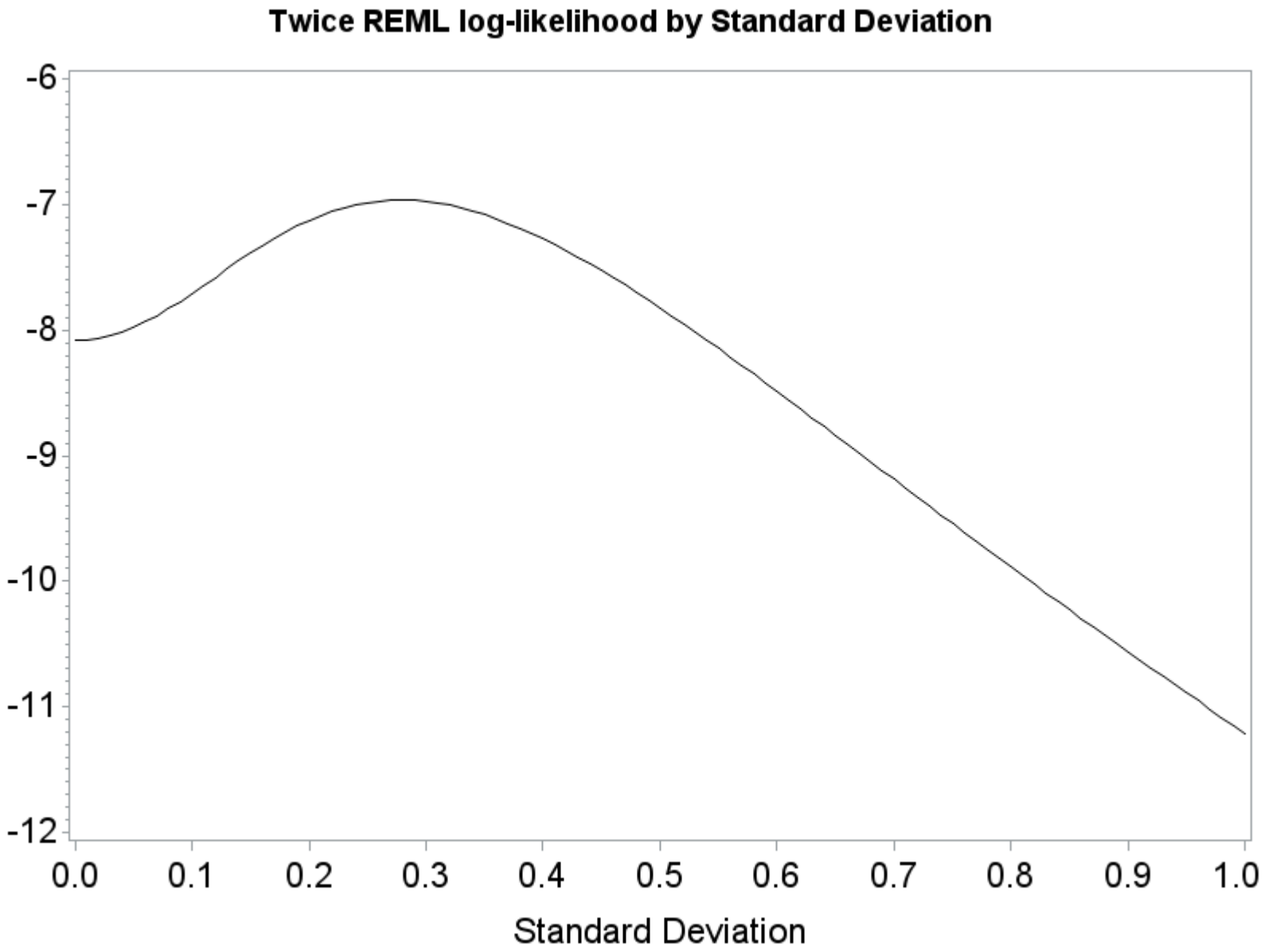
We now generate an example where the ML estimate gives a positive variance at trial level.

Quadruple the Numbers....

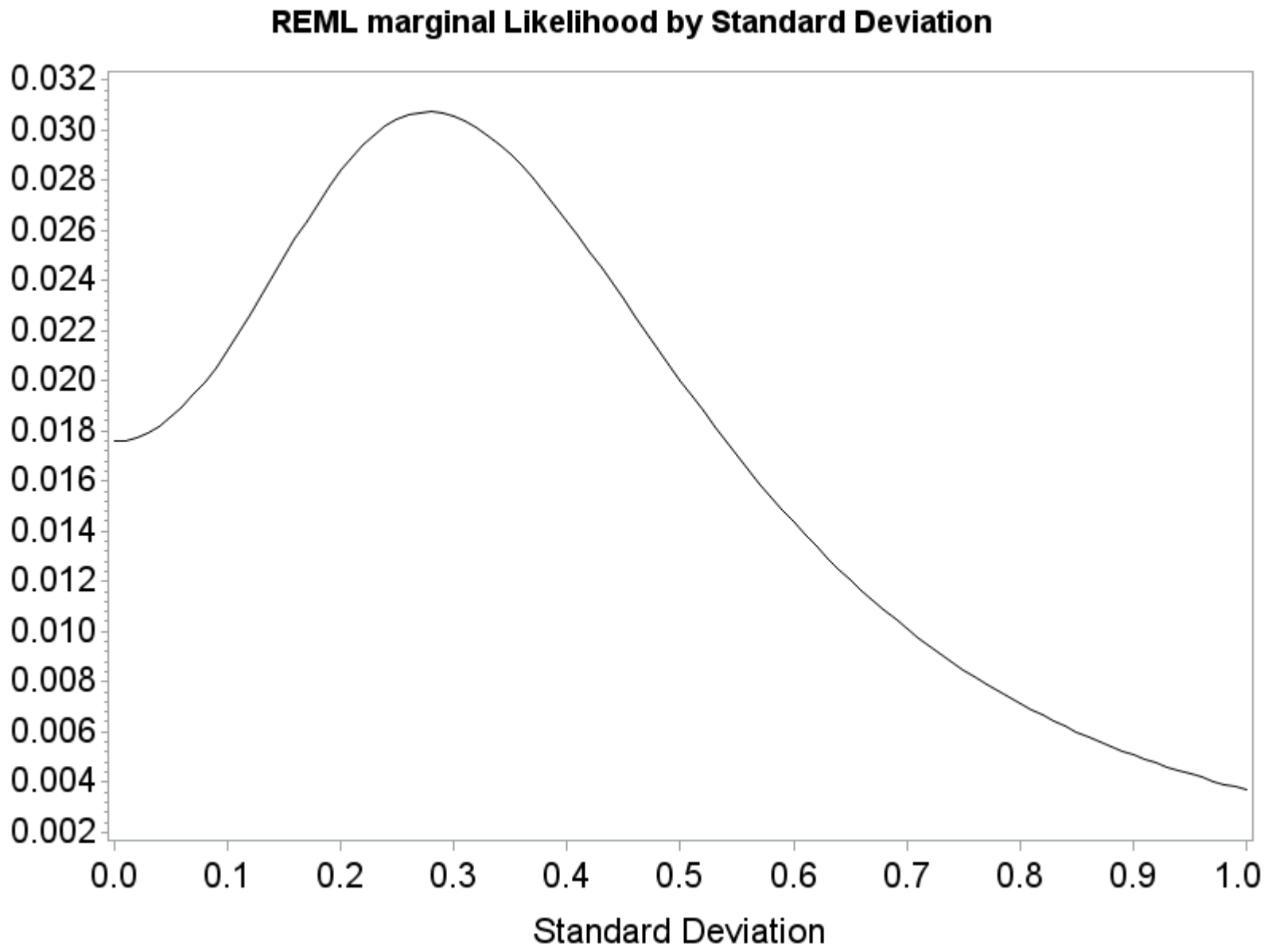
- To demonstrate the estimation of study level variability when it is positive, we modify the current data.
- Assume that each trial arm has 4 times the amount of data with the same SD.
- This reduces the variability at the within study level (due to sampling within study).
- This increases the estimated variability at the study level.

Study	Treatment	y	sd	n	Difference [Calculated]	Se(diff) [Calculated]
1	1	-1.22	3.7	4 * 54		
	3	-1.53	4.28	4 * 95	-0.31	0.668/2
2	1	-0.7	3.7	4 * 172		
	2	-2.4	3.4	4 * 173	-1.7	0.383/2
3	1	-0.3	4.4	4 * 76		
	2	-2.6	4.3	4 * 71	-2.3	0.718/2
	4	-1.2	4.3	4 * 81	-0.9	0.695/2
4	3	-0.24	3	4 * 128		
	4	-0.59	3	4 * 72	-0.35	0.442/2
5	3	-0.73	3	4 * 80		
	4	-0.18	3	4 * 46	0.55	0.555/2
6	4	-2.2	2.31	4 * 137		
	5	-2.5	2.18	4 * 131	-0.3	0.274/2
7	4	-1.8	2.48	4 * 154		
	5	-2.1	2.99	4 * 143	-0.3	0.320/2

Using quadruple data.



Using the quadruple data



Specify known variances on RANDOM

```
proc mixed data= Parkinsonstimes4;
class Study Treatment Record;
model  Y = Study Treatment  /solution  ddfm=kr ;
random SE / subject=Study*Treatment;
parms 1 1 / HOLD=(1);
lsmeans Treatment / diff=control("1");
run;
```

Specify known variances on REPEATED

```
proc mixed data=Parkinsonstimes4 ;
class Study Treatment;
model Y= Study Treatment / solution ddfm=kr;
random intercept /subject=Study*Treatment ;
parms 1 1 / hold=(2);
weight Weight;
lsmeans Treatment / diff=control("1");
run;
```

Covariance parameter estimates

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SE	Study*Treatment	1.0000
Residual		0.03876

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	Study*Treatment	0.03876
Residual		1.0000

0.03876 is variance $\sigma^2/2$.

So SD for random effect model is $\sqrt{}$(2 * 0.03876) = 0.2784

which matches the REML profile plot.

Treatment differences.

Differences of Least Squares Means

Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|
Treatment	2	1	-1.8747	0.2767	3	-6.78	0.0066
Treatment	3	1	-0.5120	0.3249	4	-1.58	0.1902
Treatment	4	1	-0.5314	0.3190	4	-1.67	0.1711
Treatment	5	1	-0.8314	0.3895	3.23	-2.13	0.1161

- **Note the very small d.f. from DDFM=KR.**
  - **This is because the random effect variance is so poorly estimated.**
- **When variance is on the boundary at zero KR does not apply and the d.f. are very large. Variance is assumed known at zero.**
- **KR is performing close to its limits. Take care when denominator degrees of freedom go lower than about 3.**
- **Note that SEDs are smaller than even fixed effects model.**
  - **Although variability pushed up to between study level, it has less impact.**

Note implication.

- If, in contrast, we overestimate the SE within trial then this will reduce the variability between studies.
  - Might be because estimates are based on covariate adjustments but SEs are not (using raw SDs say).
  - Randomization at Centre level without Centre in the model.
  - Rounding of source data.
- Although this reduces the between study variability, the overall impact is conservative, increasing the SED for indirect comparisons.

Summary: Normal data

- Fixed effects.
  - Use MIXED or GENMOD.
  - Frequentist and Bayesian effectively the same. So why bother!
  - Bayes statement on GENMOD makes Bayesian very easy.

- Random effect
  - We expect Frequentist and Bayesian to be different.
  - Frequentist easy with MIXED or GENMOD using WEIGHT.
  - Bayesian easy with the MCMC procedure.
  - Use trick of adding variability to residual rather than specify individual random effects.
  - (1) and (2) for Omega are identical for REML or Bayes with flat priors for Study fixed effect. Use whichever is easiest, usually (1).

Final message

- The “Random effects” in the casual term “random effects model” refers to the difference between treatments.
  - The treatment* Study interaction is random.
- Study is treated as a fixed effect.
  - This means all information come from with trial and is fully randomized.
- Beware any analysis (Bayesian or otherwise) where Study is treated as a random effect.

WORKSHOP 2

Workshop

- Normal data.

Scott et al

D. A. Scott, K. S. Boye, L. Timlin, J. F. Clark & J.H. Best (2013)

A network meta-analysis to compare glycaemic control in patients with type 2 diabetes treated with exenatide once weekly or liraglutide once daily in comparison with insulin glargine, exenatide twice daily or placebo.

Diabetes, Obesity and Metabolism 15: 213–223.

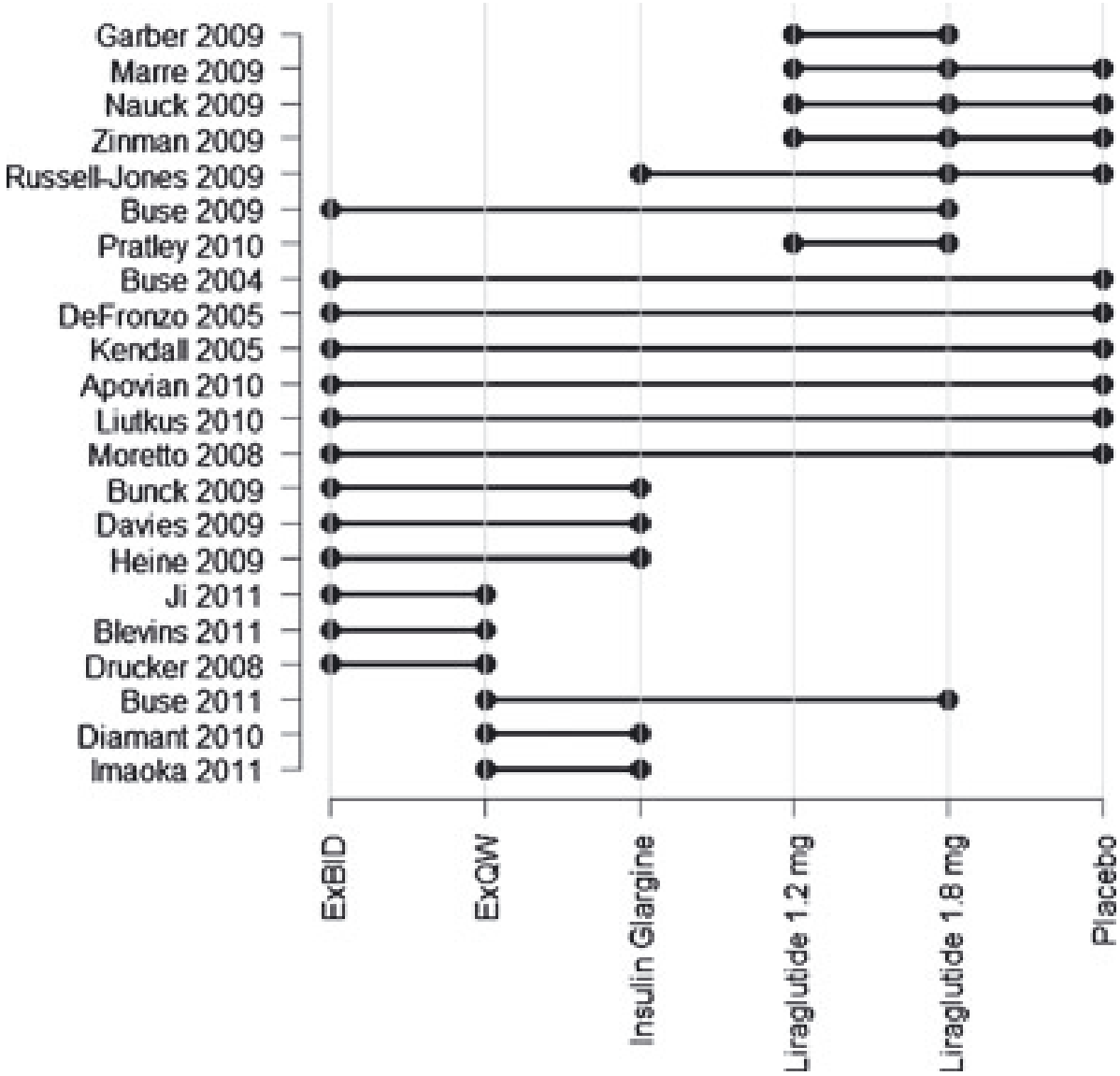
Summary

The paper's aims:

The glucagon-like peptide-1 receptor agonists (GLP-1 RAs) exenatide once weekly (ExQW) and liraglutide once daily (QD) are indicated to improve glycaemic control in patients with type 2 diabetes.

Although glycaemic control with ExQW versus liraglutide QD 1.8 mg has been directly compared, no studies have compared ExQW with liraglutide QD 1.2 mg or determined the probable relative efficacies of various injectable therapies for glycaemic control; therefore, a network meta-analysis was performed to address these questions.

The network


Garber 2009
Marre 2009
Nauck 2009
Zinman 2009
Russell-Jones 2009
Buse 2009
Pratley 2010
Buse 2004
DeFronzo 2005
Kendall 2005
Apovian 2010
Liutkus 2010
Moretto 2008
Bunck 2009
Davies 2009
Heine 2009
Ji 2011
Blevins 2011
Drucker 2008
Buse 2011
Diamant 2010
Imaoka 2011
ExBID
ExQW
Insulin Glargine
Liraglutide 1.2 mg
Liraglutide 1.8 mg
Placebo

Studies.

- 22 studies.
- 48 records (Study*Treat combinations).
- 6 treatments including Placebo

Actions

- Follow the steps in the handout.
- Program file is Workshop2.sas
- We will discuss our results at the end.

Fixed effects model

```
Title1 "Fixed effects basic model";
proc mixed data=Scott3 ;
class Study Treatment;
model Y= Study Treatment /ddf=500,500;
weight Weight;
parms 1 /hold=(1);
lsmeans Treatment / diff=control("Placebo") df=500 CL;
run;
```

Fixed effect results.

Differences of Least Squares Means										
Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Treatment	Exenatide BID	Placebo	-0.7900	0.04838	500	-16.33	<.0001	0.05	-0.8850	-0.6949
Treatment	Exenatide QW	Placebo	-1.1149	0.06084	500	-18.33	<.0001	0.05	-1.2344	-0.9954
Treatment	Insulin Glargine	Placebo	-0.8172	0.06270	500	-13.03	<.0001	0.05	-0.9404	-0.6940
Treatment	Liraglutide 1.2mg	Placebo	-1.0313	0.06926	500	-14.89	<.0001	0.05	-1.1674	-0.8952
Treatment	Liraglutide 1.8mg	Placebo	-1.2050	0.05694	500	-21.16	<.0001	0.05	-1.3169	-1.0931

Random effects model

```
Title1 "Random effects basic model";
proc mixed data=Scott3 ;
class Study Treatment;
model Y= Study Treatment /ddfm=KR;
random intercept /subject=Study*Treatment ;
weight Weight;
parms 1 1 /hold=(2);
lsmeans Treatment / CL diff=control("Placebo") L;
run;
```

Differences of Least Squares Means										
Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Treatment	Exenatide BID	Placebo	-0.7902	0.06369	21	-12.41	<.0001	0.05	-0.9226	-0.6577
Treatment	Exenatide QW	Placebo	-1.1434	0.08608	18	-13.28	<.0001	0.05	-1.3243	-0.9626
Treatment	Insulin Glargine	Placebo	-0.8241	0.08653	20.6	-9.52	<.0001	0.05	-1.0042	-0.6439
Treatment	Liraglutide 1.2mg	Placebo	-1.0372	0.09062	21	-11.44	<.0001	0.05	-1.2256	-0.8487
Treatment	Liraglutide 1.8mg	Placebo	-1.1880	0.07537	21	-15.76	<.0001	0.05	-1.3448	-1.0313

Random effects

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	study*Treatment	0.005891
Residual		1.0000

sqrt(0.005891*2) = 0.1085449

Select difference with smallest d.f.

Differences of Least Squares Means										
Effect	Treatment	_Treatment	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Treatment	Exenatide BID	Exenatide QW	0.3533	0.07047	16.1	5.01	0.0001	0.05	0.2040	0.5025
Treatment	Exenatide BID	Insulin Glargine	0.03387	0.07204	17.6	0.47	0.6440	0.05	-0.1177	0.1854
Treatment	Exenatide BID	Liraglutide 1.2mg	0.2470	0.09730	21	2.54	0.0191	0.05	0.04462	0.4493
Treatment	Exenatide BID	Liraglutide 1.8mg	0.3978	0.07740	18.9	5.14	<.0001	0.05	0.2358	0.5599
Treatment	Exenatide BID	Placebo	-0.7902	0.06369	21	-12.41	<.0001	0.05	-0.9226	-0.6577
Treatment	Exenatide QW	Insulin Glargine	-0.3194	0.07365	12.4	-4.34	0.0009	0.05	-0.4793	-0.1595
Treatment	Exenatide QW	Liraglutide 1.2mg	-0.1063	0.1060	16.9	-1.00	0.3300	0.05	-0.3300	0.1174
Treatment	Exenatide QW	Liraglutide 1.8mg	0.04458	0.08492	13.2	0.53	0.6083	0.05	-0.1386	0.2278
Treatment	Exenatide QW	Placebo	-1.1434	0.08608	18	-13.28	<.0001	0.05	-1.3243	-0.9626
Treatment	Insulin Glargine	Liraglutide 1.2mg	0.2131	0.1090	19.2	1.96	0.0653	0.05	-0.01485	0.4411
Treatment	Insulin Glargine	Liraglutide 1.8mg	0.3640	0.08975	16.2	4.06	0.0009	0.05	0.1739	0.5540
Treatment	Insulin Glargine	Placebo	-0.8241	0.08653	20.6	-9.52	<.0001	0.05	-1.0042	-0.6439
Treatment	Liraglutide 1.2mg	Liraglutide 1.8mg	0.1509	0.07331	21	2.06	0.0522	0.05	-0.00160	0.3033
Treatment	Liraglutide 1.2mg	Placebo	-1.0372	0.09062	21	-11.44	<.0001	0.05	-1.2256	-0.8487
Treatment	Liraglutide 1.8mg	Placebo	-1.1880	0.07537	21	-15.76	<.0001	0.05	-1.3448	-1.0313

Why?

- Strange as it is a comparison with lots of direct links, so you might expect that the SD will depend less on the RE.

- But these are studies with large numbers of subject leading to small SED, where the between study variation (RE) starts to dominate.

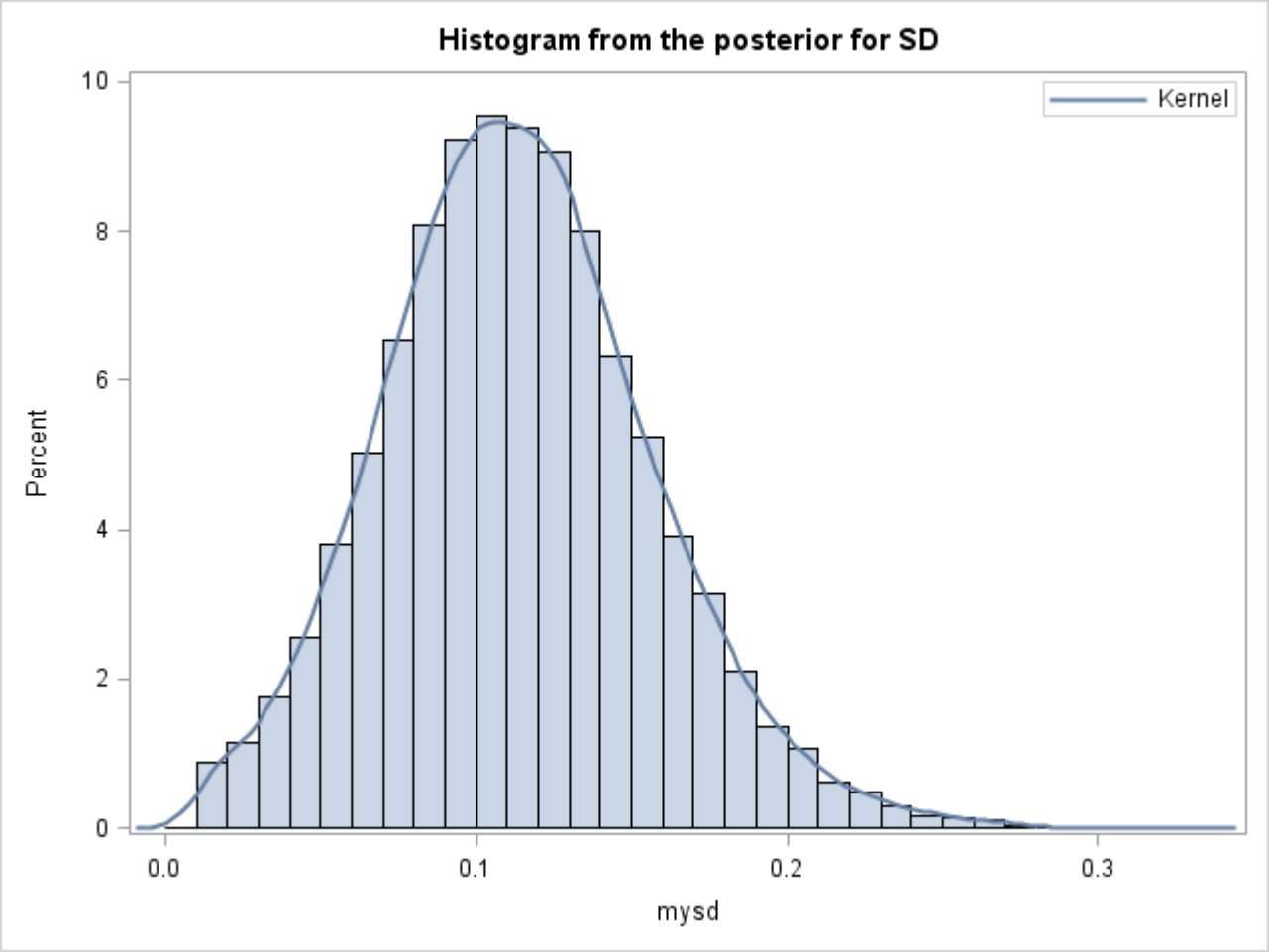
- Here the frequentist approach is helping with interpretation.

The MCMC code.

- Takes much longer to run.
- Using the trick of not defining random effects directly makes mixing better.
- The SAS 9.3 code is easier.

- The RE SD is well estimated in this example so there is no difference between frequentist and Bayesian.

The RE SD is well estimated.



NMA METHODOLOGY (CONT.)

BINOMIAL DATA

CASE STUDY WITH BINARY DATA

Introduction

- Yes / No outcome.
- Often safety endpoints.
- Most common form of data for indirect comparisons.
- Rare events may introduce difficulties.

Example 1:
Pagliaro et al (1992),
Annals of internal Medicine,117,59-70.

As used in

- Higgins & Whitehead (1996) *Borrowing strength from external trials in meta-analysis*. Statistics in Medicine, 15: 2733–2749.
- Whitehead (2002) *Meta-analysis of controlled clinical trials*. Wiley.
- Lu & Ades (2004) *Combination of direct and indirect evidence in mixed treatment comparisons.* Statistics in Medicine, 23, 3105-3124.
- Jones, Roger, Lane, Lawton, Fletcher, Cappelleri et al. (2011) *Statistical approaches for conducting network meta-analysis in drug development.* Pharmaceutical Statistics, 10, 523-531

Pagliaro et al.

- 26 studies
- Prevention of first bleeding in cirrhosis.
- Compare either two or three nonsurgical treatments (A, B and C).
  - A was the use of Beta-blockers
  - B was Sclerotherapy
  - C was a Control treatment.

- Data expressed as R events out of N subjects.

- Studies 1 and 2 have three arms
- All the rest compare A to C, or compare B to C

APPENDIX A. DATASET

Table A.1. Data from [3]. Table arranged to display studies directly comparing A and C on the left side and studies directly comparing B and C on the right side.

Set 1					Set 2				
	A		C			B		C	
Study Label	*r*	*n*	*r*	*n*	Study Label	*r*	*n*	*r*	*n*
1	2	43	13	41	1	9	42	13	41
2	12	68	13	72	2	13	73	13	72
3	4	20	4	16	10	4	18	0	19
4	20	116	30	111	11	3	35	22	36
5	1	30	11	49	12	5	56	30	53
6	7	53	10	53	13	5	16	6	18
7	18	85	31	89	14	3	23	9	22
8	2	51	11	51	15	11	49	31	46
9	8	23	2	25	16	19	53	9	60
					17	17	53	26	60
					18	10	71	29	69
					19	12	41	14	41
					20	0	21	3	20
					21	13	33	14	35
					22	31	143	23	138
					23	20	55	19	51
					24	3	13	12	16
					25	3	21	5	28
					26	6	22	2	24

- Studies 1 and 2 appear in both set 1 and 2 (3 arm studies)
- Quite large differences in size of study (roughly equal sized arms within).

```
Data Betablock1;
Input  Study     Ar          An          Br          Bn          Cr          Cn          ;
datalines;
1           2           43          9           42          13          41
2           12          68          13          73          13          72
3           4           20          0           0           4           16
4           20          116         0           0           30          111
5           1           30          0           0           11          49
6           7           53          0           0           10          53
7           18          85          0           0           31          89
8           2           51          0           0           11          51
9           8           23          0           0           2           25
10          0           0           4           18          0           19
11          0           0           3           35          22          36
12          0           0           5           56          30          53
13          0           0           5           16          6           18
14          0           0           3           23          9           22
15          0           0           11          49          31          46
16          0           0           19          53          9           60
17          0           0           17          53          26          60
18          0           0           10          71          29          69
19          0           0           12          41          14          41
20          0           0           0           21          3           20
21          0           0           13          33          14          35
22          0           0           31          143         23          138
23          0           0           20          55          19          51
24          0           0           3           13          12          16
25          0           0           3           21          5           28
26          0           0           6           22          2           24
; run;
```

Make data vertical (one row per arm)

```
data betablock3;
set betablock1 end=myend;
by study;
length Trt $ 8;
retain Nreff 0 NTrtlev 0 Nstudy 0;
drop ar an br bn cr cn Ntrtlev Record Nreff
     NStudy;
retain record 0;
Narm= (ar+an >0) + (br+bn >0) + (cr+cn >0) ;
arm=0;
index=0;
if ar+an >0 then do;
     arm=arm+1;
     r=ar;
     n=an;
     Trt="A";
     ITrt=1;
     record=record+1;
     index=index+1;
     output;
end;
if br+bn >0 then do;
          arm=arm+1;
          r=br;
          n=bn;
          Trt="B";
          ITrt=2;
          index=index+1;
          record=record+1;
          output;
end;
if cr+cn >0 then do;
          arm=arm+1;
          r=cr;
          n=cn;
          Trt="C";
          ITrt=3;
          index=index+1;
          record=record+1;
          output;
end;
Nstudy=Nstudy+1;
Nreff=Nreff+Narm-1;
NTrtlev=max(Ntrtlev,narm);
if myend then do;
          call symput("Nrec",record);
          call symput("Nreff",Nreff);
          call symput("NTrtlev",NTrtlev);
          call symput("NStudy",NStudy);
end;
run;
```

Pagliaro in Vertical format

Study	Narm	Index	R	N	Trt
1	3	1	2	43	1
1	3	2	9	42	2
1	3	3	13	41	3
2	3	1	12	68	1
2	3	2	13	73	2
2	3	3	13	72	3
3	2	1	4	20	1
3	2	2	4	16	3
4	2	1	20	116	1
4	2	2	30	111	3
5	2	1	1	30	1
5	2	2	11	49	3
6	2	1	7	53	1
6	2	2	10	53	3

Statistical overview

- The problem falls within the class of generalized linear models.
- But we may need to add random effects.
- Generalized mixed models can be fitted and interpreted in two ways:
  - Marginal models [e.g. REPEATED statement in proc GENMOD.]
  - Subject-specific models. Here the model is defined conditional upon the random subject effect.
- We follow the second route.

Fixed effects – Odds ratios

- Logistic regression.
- Logit link function and Binomial distribution.

- Issues
  - Maximum likelihood approach relies on asymptotic results.
  - Use EXACT approach with very rare events.
  - Often interpreted in terms of LogOdds ratios (LOR).

$$X\beta = log\left[\frac{R/N}{1 - R/N}\right]$$

$$LOR = log\left[\frac{R_1(N_2 - R_2)}{(N_1 - R_1)R_2}\right] = X\beta_1 - X\beta_2$$

Fixed effects – Relative risk

- Log link function and Binomial distribution.

- Issues
  - Modelled probability can in theory go > 1
  - But usually use with Yes/No arranged to give small probabilities.

$$X\beta = log[R/N]$$

$$Log(RR) = log\left[\frac{R_1/N_1}{R_2/N_2}\right] = X\beta_1 - X\beta_2$$

Generalized linear model (within study)

- Both assume a Binomial distribution for R / N.
- Logistic link or Log link.
  - We will later see the use of the complementary log-log as link function.
- For low rate events such as rare adverse events, both behave similarly.

- Many people in this area want to contrast *direct information* from *indirect information*.

- For this purpose we will temporarily analyse studies 1 and 2 separately from the rest.
  - Note that as there are no estimated variance parameters, this is possible – no worry about sharing of parameters.

Ignoring the three arm studies
Logistic fixed effects model (no RE)

```
proc genmod data=Betablock2 desc;
where Study notin(1,2);
class Study Trt;
model R/N = Study Trt /link=logit dist=bin type1;
lsmeans Trt /diff exp cl;
run;
```

Logistic removing studies 1 and 2

Differences of Trt Least Squares Means											
Trt	_Trt	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
1	2	0.004680	0.2208	0.02	0.9831	0.05	-0.4280	0.4374	1.0047	0.6518	1.5486
1	3	-0.6033	0.1847	-3.27	0.0011	0.05	-0.9654	-0.2413	0.5470	0.3808	0.7856
2	3	-0.6080	0.1209	-5.03	<.0001	0.05	-0.8450	-0.3710	0.5444	0.4296	0.6900

Direct comparisons A-C and B-C

Indirect comparison A-B

Fixed effects analysis

Table II. Estimates (standard errors) on the log odds ratios scale for direct and indirect treatment comparisons.

Comparison	Direct estimate studies	Indirect estimate studies 3-26	Direct estimate studies 1-2	MTC estimate all studies	Bayesian MTC all studies
A – *C*	–0.603(0.185)	—	–0.730(0.363)	–0.670(0.161)	–0.679(0.161)
B – *C*	–0.608(0.121)	—	–0.234(0.326)	–0.553(0.113)	–0.559(0.113)
A – *B*	—	0.005(0.221)	–0.496(0.371)	–0.117(0.189)	–0.120(0.190)

MTC = mixed treatment comparison.

- Traditionally interest has focused on “direct” and “indirect” sources of information.
- Note how A-B contrast has larger SED as it is mostly estimated indirectly.
- Bayesian results can be obtained using BAYES statement using GENMOD;

Logistic, studies 1 and 2 only

Differences of Trt Least Squares Means											
Trt	_Trt	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
1	2	-0.4963	0.3715	-1.34	0.1816	0.05	-1.2245	0.2318	0.6088	0.2939	1.2609
1	3	-0.7304	0.3631	-2.01	0.0442	0.05	-1.4420	-0.01883	0.4817	0.2365	0.9813
2	3	-0.2341	0.3259	-0.72	0.4726	0.05	-0.8728	0.4047	0.7913	0.4178	1.4988

Logistic, all studies together

Differences of Trt Least Squares Means											
Trt	_Trt	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
1	2	-0.1172	0.1892	-0.62	0.5357	0.05	-0.4879	0.2536	0.8894	0.6139	1.2887
1	3	-0.6700	0.1608	-4.17	<.0001	0.05	-0.9852	-0.3548	0.5117	0.3734	0.7013
2	3	-0.5528	0.1125	-4.91	<.0001	0.05	-0.7733	-0.3324	0.5753	0.4615	0.7172

bayes diag=all statistics=all NMC=10000 seed=12345;

Sample Differences of Trt Least Squares Means																	
Trt	_Trt	N	Estimate	Standard Deviation	Percentiles			Alpha	Lower HPD	Upper HPD	Exponentiated	Standard Error of Exponentiated	Percentiles for Exponentiated			Lower HPD of Exponentiated	Upper HPD of Exponentiated
					25th	50th	75th						25th	50th	75th		
1	2	10000	-0.1239	0.1919	-0.2518	-0.1212	0.00519	0.05	-0.4904	0.2489	0.8998	0.173030	0.7774	0.8859	1.0052	0.5683	1.2223
1	3	10000	-0.6788	0.1617	-0.7883	-0.6813	-0.5730	0.05	-0.9873	-0.3678	0.5139	0.083560	0.4546	0.5059	0.5638	0.3726	0.6923
2	3	10000	-0.5549	0.1145	-0.6265	-0.5547	-0.4816	0.05	-0.7878	-0.3457	0.5779	0.066140	0.5344	0.5743	0.6178	0.4548	0.7077

Using MCMC procedure SAS 9.3

```
proc mcmc data=betablock3  nmc=200000 seed=246810;
random Studyeffect ~general(0) subject=Study init=(0) ;
random Treat ~general(0) subject=Trt  init=(0) zero=last monitor=(Treat);
Mu=  Studyeffect + Treat ;
P=1-(1/(1+exp(mu)));
model R ~ binomial(n=N, p=P);
run;
```

- Zero=Last makes contrasts compare to Control treatment C.
- Mixes so well no need for thinning.

Proc MCMC results

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Treat_A	200000	-0.6798	0.1622	-0.7878	-0.6792	-0.5702
Treat_B	200000	-0.5597	0.1131	-0.6356	-0.5597	-0.4830

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Treat_A	0.050	-1.0024	-0.3638	-1.0001	-0.3617
Treat_B	0.050	-0.7816	-0.3371	-0.7821	-0.3379

Fixed effects analysis

Table II. Estimates (standard errors) on the log odds ratios scale for direct and indirect treatment comparisons.

Comparison	Direct estimate studies	Indirect estimate studies 3-26	Direct estimate studies 1-2	MTC estimate all studies	Bayesian MTC all studies
A – *C*	–0.603(0.185)	—	–0.730(0.363)	–0.670(0.161)	–0.679(0.161)
B – *C*	–0.608(0.121)	—	–0.234(0.326)	–0.553(0.113)	–0.559(0.113)
A – *B*	—	0.005(0.221)	–0.496(0.371)	–0.117(0.189)	–0.120(0.190)

MTC = mixed treatment comparison.

- Bayesian analysis gives very similar results to maximum likelihood (ML). As expected!
- MTC summarises combination of direct and indirect comparisons.

Random effects Binary Data

- This is the truly classic problem.

Random effects model

- Study i and Arm k, with Treatment t(i,k)

$$R_{ik} \mid \eta_{ik} \sim Bin(N_{ik}, P_{ik})$$

$$log\left(\frac{P_{ik}}{1-P_{ik}}\right) = \mu_i + \delta_{t(i,k)} + \eta_{ik}$$

where η_{ik} has zero mean, independent between studies with

$Cov(\eta_{ik}, \eta_{ih}) = \omega_{kh}$

Note Study and Treatment main effects remain as FIXED.

The Statistical model
Omega as before.

- Symmetry is assumed leading to two possible options

where the i'th study has m_i arms (Ω is m_i by m_i).

1) $\omega_{kk} = \sigma^2/2$ and $\omega_{kh} = 0$ if $k \neq h$.

this is a simple diagonal matrix.

2) $\omega_{kk} = \sigma^2 - \sigma^2/2m_i$

and $\omega_{kh} = -\sigma^2/2m_i$ if $k \neq h$.

The two forms for Omega.

- For 2 and 3 arm trials

(1)

$$\begin{bmatrix} \sigma^2/2 & 0 \\ 0 & \sigma^2/2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2/2 & 0 & 0 \\ 0 & \sigma^2/2 & 0 \\ 0 & 0 & \sigma^2/2 \end{bmatrix}$$

(2)

$$\begin{bmatrix} \sigma^2/4 & -\sigma^2/4 \\ -\sigma^2/4 & \sigma^2/4 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2/3 & -\sigma^2/6 & -\sigma^2/6 \\ -\sigma^2/6 & \sigma^2/3 & -\sigma^2/6 \\ -\sigma^2/6 & -\sigma^2/6 & \sigma^2/3 \end{bmatrix}$$

Generalized Mixed Models (GLMM)

- In the Normal case the marginal distribution of Y integrated over the random effect is known and Normal.
  - But with GLMM Frequentist has to use an approximation or use numerical integration.
  - Bayesian can use MCMC with random effects as variables in the hierarchic model.

- Up until recently these were fitted using Pseudo-Quasi-likelihood (PQL) algorithms (approximation).
  - Does not work well with Binary data and logistic link.
  - Useful with Binomial data as long as Normal approximation is OK.
  - Can use REML within PQL to allow for estimation of linear model effects in estimating variance components.

Generalized Mixed Models (GLMM)

- Favoured approach in GLMM circles is now to use some form of Gaussian Quadrature (Numerical integration), or Laplace approximation.
  - Used to require NLMIXED, but now can use GLIMMIX.
  - But no current equivalent to REML (an issue here).

- So now we move to using the GLIMMIX procedure.
  - For PQL with REML
  - For Gaussian quadrature.

GLIMMIX – Fixed effects (again)

```
proc glimmix data=Betablock2 ;
class Study Trt;
model R/N = Study Trt /link=logit dist=bin ddfm=none;
lsmeans Trt /diff cl oddsratios;
run;
```

- Logit link and Binomial distribution.
- Use ddfm=none to get equivalent results to GENMOD.

Fixed effects logistic using GLIMMIX

The GLIMMIX Procedure

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1172	0.1892	Infty	-0.62	0.5357	0.05	-0.4879	0.2536	0.889	0.614	1.289
1	3	-0.6700	0.1608	Infty	-4.17	<.0001	0.05	-0.9852	-0.3548	0.512	0.373	0.701
2	3	-0.5528	0.1125	Infty	-4.91	<.0001	0.05	-0.7733	-0.3324	0.575	0.461	0.717

Results the same as before.

So add the random effect as (1)

```
proc glimmix data=Betablock2 method=RSPL;
class Study Trt;
model R/N = Trt Study/link=logit dist=bin ddfm=none;
random Trt / subject=Study;
lsmeans Trt /diff cl oddsratios;
run;
```

- This uses PQL algorithm with REML (actually the default).

PQL with Omega (1)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Trt	Study	0.5813	0.2286

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1620	0.4673	Infty	-0.35	0.7289	0.05	-1.0778	0.7539	0.850	0.340	2.125
1	3	-0.7364	0.4014	Infty	-1.83	0.0666	0.05	-1.5230	0.05034	0.479	0.218	1.052
2	3	-0.5744	0.2814	Infty	-2.04	0.0413	0.05	-1.1260	-0.02279	0.563	0.324	0.977

- Note variance is half the variance of treatment difference σ^2 as specified before and as in Jones et al. So σ^2 =1.1626.
- This is slightly less than then median from the Bayesian posterior for SD we get later (1.4).

Bayes solutions from Jones et al.

Table IV. Estimates: mean (standard deviation) of posterior distribution for log-odds and median for variance σ^2, obtained using WinBUGS and the MCMC procedure in SAS.

	Whitehead (2002) code	Bristol code	MCMC in *SAS*®
$A - C$	−0.784(0.442)	−0.784(0.455)	−0.783(0.458)
$B - C$	−0.599(0.312)	−0.598(0.319)	−0.599(0.322)
$A - B$	−0.185(0.515)	−0.185(0.531)	−0.184(0.535)
Variance (σ^2)	1.330	1.453	1.458

MCMC, Markov Chain Monte Carlo.

B. Jones et al.

Gaussian Quadrature with Omega (1)

```
proc glimmix data=Betablock2 method=QUAD;
class Study Trt;
model R/N = Trt Study/link=logit dist=bin ddfm=none;
random Trt / subject=Study;
lsmeans Trt /diff cl oddsratios;
run;
```

- Use of NLMIXED gives identical results.

Gaussian Quadrature with Omega (1)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Trt	Study	0.1613	0.06505

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1316	0.3021	Infty	-0.44	0.6632	0.05	-0.7237	0.4606	0.877	0.485	1.585
1	3	-0.7171	0.2596	Infty	-2.76	0.0057	0.05	-1.2259	-0.2082	0.488	0.293	0.812
2	3	-0.5855	0.1814	Infty	-3.23	0.0012	0.05	-0.9411	-0.2300	0.557	0.390	0.795

- Estimate of variability much lower.
- Implies Standard errors much smaller. Is this due to using ML?
- Note variance is half the variance of treatment difference σ^2 as specified before and as in Jones et al. So σ^2 =0.3226, compared to 1.16 before.

proc glimmix data=Betablock2 Method=MSPL; Maximum Likelihood with PQL. Omega (1)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Trt	Study	0.1529	0.06081

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1285	0.2963	Infty	-0.43	0.6645	0.05	-0.7093	0.4523	0.879	0.492	1.572
1	3	-0.7037	0.2545	Infty	-2.77	0.0057	0.05	-1.2025	-0.2050	0.495	0.300	0.815
2	3	-0.5752	0.1780	Infty	-3.23	0.0012	0.05	-0.9240	-0.2264	0.563	0.397	0.797

- With ML rather than REML the PQL variance estimate is slightly smaller than with Gaussian quadrature, which is expected from theory.
- σ^2 =0.3058.
- This should be (partly) fixed by using Omega (2) rather than Omega (1).

Omega (2) with same trick as for Normal

```
data jr;
set Betablock2;
array x[3] x1-x3;
do i=1 to 3;
      if i<= narm then x[i]=sqrt(0.5)*((i=index)-1/narm);
      else x[i]=0;
end;
run;

proc glimmix data=jr method=QUAD;
class Study Trt ;
model R/N = Trt Study/link=logit dist=bin ddfm=none;
random X1 X2 X3 / subject=Study type=TOEP(1) ;
lsmeans Trt /diff cl oddsratios;
run;
```

Gaussian Quadrature with Omega (2)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	Study	1.0246	0.4021

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1590	0.4454	Infty	-0.36	0.7211	0.05	-1.0320	0.7140	0.853	0.356	2.042
1	3	-0.7402	0.3827	Infty	-1.93	0.0531	0.05	-1.4902	0.009772	0.477	0.225	1.010
2	3	-0.5812	0.2682	Infty	-2.17	0.0302	0.05	-1.1068	-0.05563	0.559	0.331	0.946

- This is the result in Table III of Jones et al.
- This is true estimate for σ^2 (does not need doubling).

REML PQL with Omega (2)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	Study	1.1626	0.4572

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1620	0.4673	Infty	-0.35	0.7289	0.05	-1.0778	0.7539	0.850	0.340	2.125
1	3	-0.7364	0.4014	Infty	-1.83	0.0666	0.05	-1.5230	0.05034	0.479	0.218	1.052
2	3	-0.5744	0.2814	Infty	-2.04	0.0413	0.05	-1.1260	-0.02279	0.563	0.324	0.977

- This σ^2 is larger than the one for Gaussian Quadrature as REML corrects for aliasing of random effect with Study fixed effect but **also with Treatment fixed effect**.
- Note that this is the same as σ^2 =1.162 from REML PQL with Omega (1).
- REML has meant that the aliasing of fixed and random effects is automatically handled.

ML PQL (MSPL) with Omega (2)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	Study	0.9949	0.3886

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1575	0.4397	Infty	-0.36	0.7202	0.05	-1.0194	0.7043	0.854	0.361	2.022
1	3	-0.7322	0.3777	Infty	-1.94	0.0526	0.05	-1.4726	0.008138	0.481	0.229	1.008
2	3	-0.5747	0.2647	Infty	-2.17	0.0300	0.05	-1.0936	-0.05580	0.563	0.335	0.946

- This σ^2 is slightly less than that for Gaussian Quadrature with Omega (2) where σ^2 =1.0246 as PQL slightly underestimates variance.
- Gaussian Quadrature with full REML-like properties is perhaps something for the future, ...

Frequentist Summary so far

- Well known that with Binomial data where counts are big enough for Normal approximation to hold, then PQL with REML is a safe approach.

- Improvement of Guassian Quadrature (GQ) over PQL in terms of approximation to likelihood, is offset by bias from fixed effect parameters (no REML).

- In this example PQL behaves much better, even when using Omega (2) approach than GQ.

For the future ... Use (I-X(X`X)$^{-1}$X`)Z instead of Z in G. Quadrature approach.

```
* Build the X and Z matrices; * Results not saved;
proc glimmix data=newdata OUTDESIGN=Fred;;
class Study Trt;
model R/N = Trt Study/link=logit dist=bin ddfm=none ;
random Trt*study;
run;

proc iml;
use x(keep=_x1-_x30);    read all into x;
use z(keep=_z1-_z54);   read all into z;
xx=x`*x;
z=z-x*ginv(xx)*x`*z;
Create Newz from z; append from z;
quit;

* Need to give the random statement a subject= variable, and will not accept INTERCEPT;
data newdata;
merge Betablock2  Newz;
James=1;
run;
```

Need to use Laplace approximation rather than full G. Quadrature.

```
proc glimmix data=newdata method=LAPLACE;
class Study Trt;
model R/N = Trt Study/link=logit dist=bin ddfm=none ;
random col1-col54 /subject=james type=toep(1);
run;
```

The GLIMMIX Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	james	0.5700	0.2266

This is the same as σ^2 =1.14, which is close to RSPL value.

- This approach is not published.

- But might be useful with rare events where PQL approximation may not hold. [But could go Bayesian, which is better documented.]

Should I use DDFM=KR?

- PQL linearises the problem so KR is used on linearised problem. Allows for fixed effect parameters in estimating SD.
- Will increase SEs and add d.f. for use with t for CI.s.
- Useful for indicating when SD is not well estimated.

Differences of Trt Least Squares Means

Trt	_Trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
1	2	-0.1620	0.4682	21.67	-0.35	0.7327	0.05	-1.1338	0.8099	0.850	0.322	2.248
1	3	-0.7364	0.4021	21.82	-1.83	0.0808	0.05	-1.5707	0.09804	0.479	0.208	1.103
2	3	-0.5744	0.2821	21.2	-2.04	0.0544	0.05	-1.1607	0.01193	0.563	0.313	1.012

was

1	2	-0.1620	0.4673	Infty	-0.35	0.7289	0.05	-1.0778	0.7539	0.850	0.340	2.125
1	3	-0.7364	0.4014	Infty	-1.83	0.0666	0.05	-1.5230	0.05034	0.479	0.218	1.052
2	3	-0.5744	0.2814	Infty	-2.04	0.0413	0.05	-1.1260	-0.02279	0.563	0.324	0.977

Summary Random effects Frequentist

- Use of Gaussian Quadrature with Omega (2) is recommended in the Jones et al paper.

- We now suggest that method=RSPL will behave better and is much easier as we can use OMEGA(1).
  - But beware when the event rate is very small, R=0,1, 2 only or when N is very small (<10 say).
  - I have no evidence that DDFM=KR is dangerous, and should allow for better estimation of SD, which is often a Bayesian's argument against this approach.

The Bayesian solution

- Nearly all the work in this area has been done using Winbugs.

- Here we show how to fit the same models using the MCMC procedure.

Winbugs

- NICE web site includes code from Bristol group.
- Messy and difficult to read
  - Expressed in terms of differences to some overall reference arm.
  - When overall reference does not appear in a study then have a “local” reference.
  - Statisticians do not need this as they simply have several fixed treatment effects with an arbitrary constraint (usually that one parameter is zero).
- Much simpler Winbugs code is possible when using “flat” priors.

Issues with MCMC procedure

- SAS 9.3 (SAS/Stat 12)
  - Take advantage of the RANDOM statement.

- Both versions
  - Use a fast machine, but especially for 9.2.

Ideas the same as for Normal, except need explicit random effects and logistic link and Binomial error.

P=1-(1/(1+exp(mu)));

model R ~ binomial(n=N, p=P);

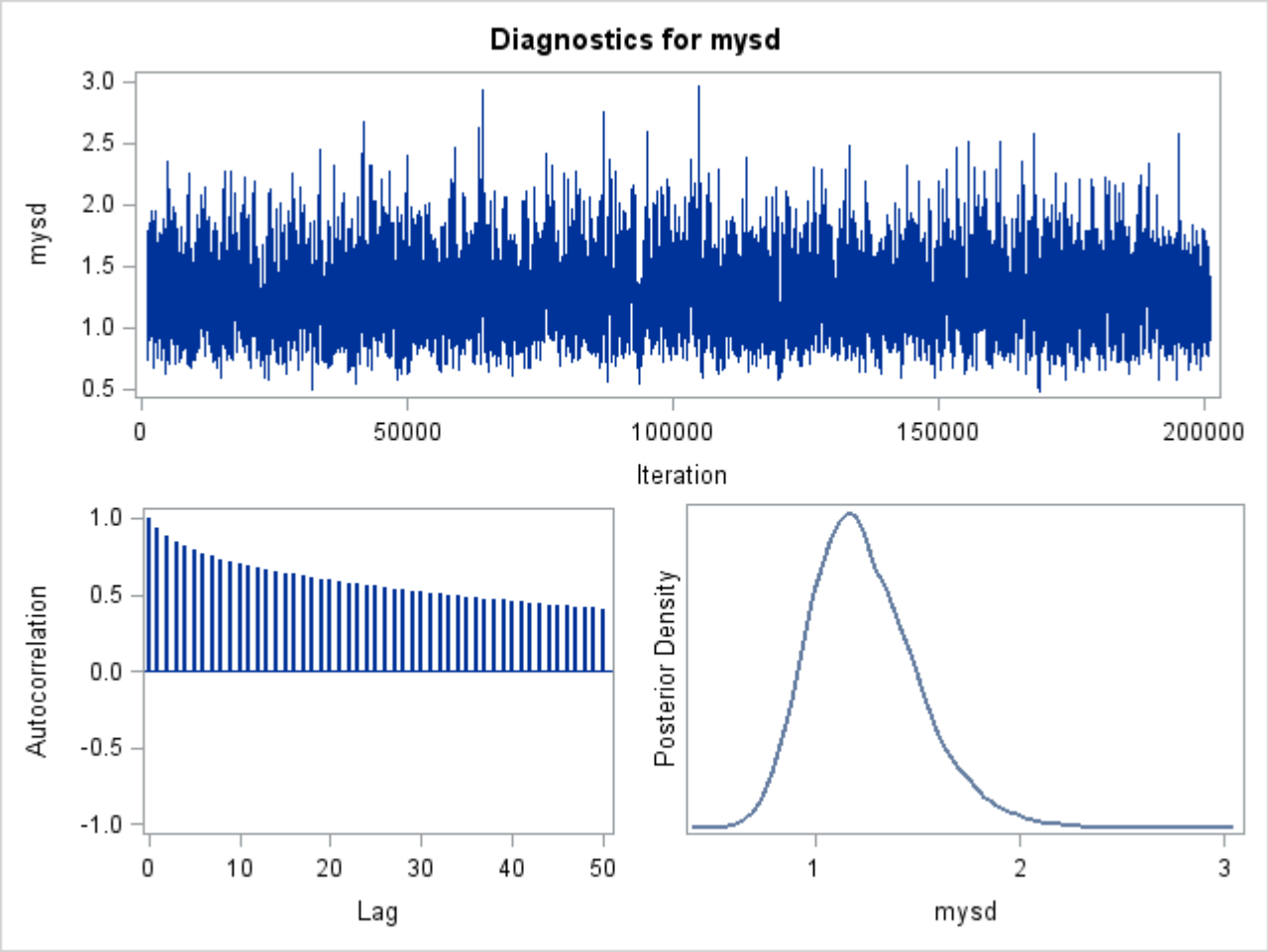
Random effects model (SAS 9.3)

```
proc mcmc data=betablock3  nmc=200000 seed=246810;
random Studyeffect ~general(0) subject=Study init=(0) ;
random Treat ~general(0) subject=Trt  init=(0) zero=last monitor=(Treat);
parms logsd 0;
prior logsd ~ general(logsd,lower=log(0.01) upper=log(5));
mysd=exp(logsd);
random RE ~normal(0,sd=mysd/sqrt(2)) subject=_OBS_ init=(0);
Mu=  Studyeffect + Treat +RE;
P=1-(1/(1+exp(mu)));
model R ~ binomial(n=N, p=P);
run;
```

Random effects solution (MCMC 9.3)

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
logsd	200000	0.1973	0.2093	0.0572	0.1960	0.3393
Treat_A	200000	-0.7604	0.4706	-1.0645	-0.7537	-0.4527
Treat_B	200000	-0.5773	0.3203	-0.7865	-0.5782	-0.3684

Diagnostics



Bayesian indirect comparison models

- If we accept flat priors on the linear predictor scale for fixed effects of treatment and study then Omega (1) and Omega (2) are identical.
  - Use Omega (1) as it is much easier.

- I do not see any reason for
  - Informative priors for study
  - Study as a Random effect.

  ... but if you do then Omega(2) is most likely necessary.

  So use the following sets of code and amend priors.

```
proc mcmc data=betablock3  nmc=200000 thin=20 seed=246810 monitor=(mysd);
random Studyeffect ~general(0) subject=Study init=(0) ;
random Treat ~general(0) subject=Trt  init=(0) zero=last monitor=(Treat);
parms logsd 0;
prior logsd ~ general(logsd,lower=log(0.01) upper=log(5));
mysd=exp(logsd);
array zero[3] (0,0,0);
array RE[3];
random RE ~mvnar(zero,sd=mysd/sqrt(2),0) subject=study;
sum=0;
do i=1 to narm;
     sum=sum+RE[i]*((i=index) - (1/narm) ) / sqrt(2);  ← Weighted sum like Yesterday.
end;
Mu=  Studyeffect + Treat +  sum;
P=1-(1/(1+exp(mu)));
model R ~ binomial(n=N, p=P);
run;
```

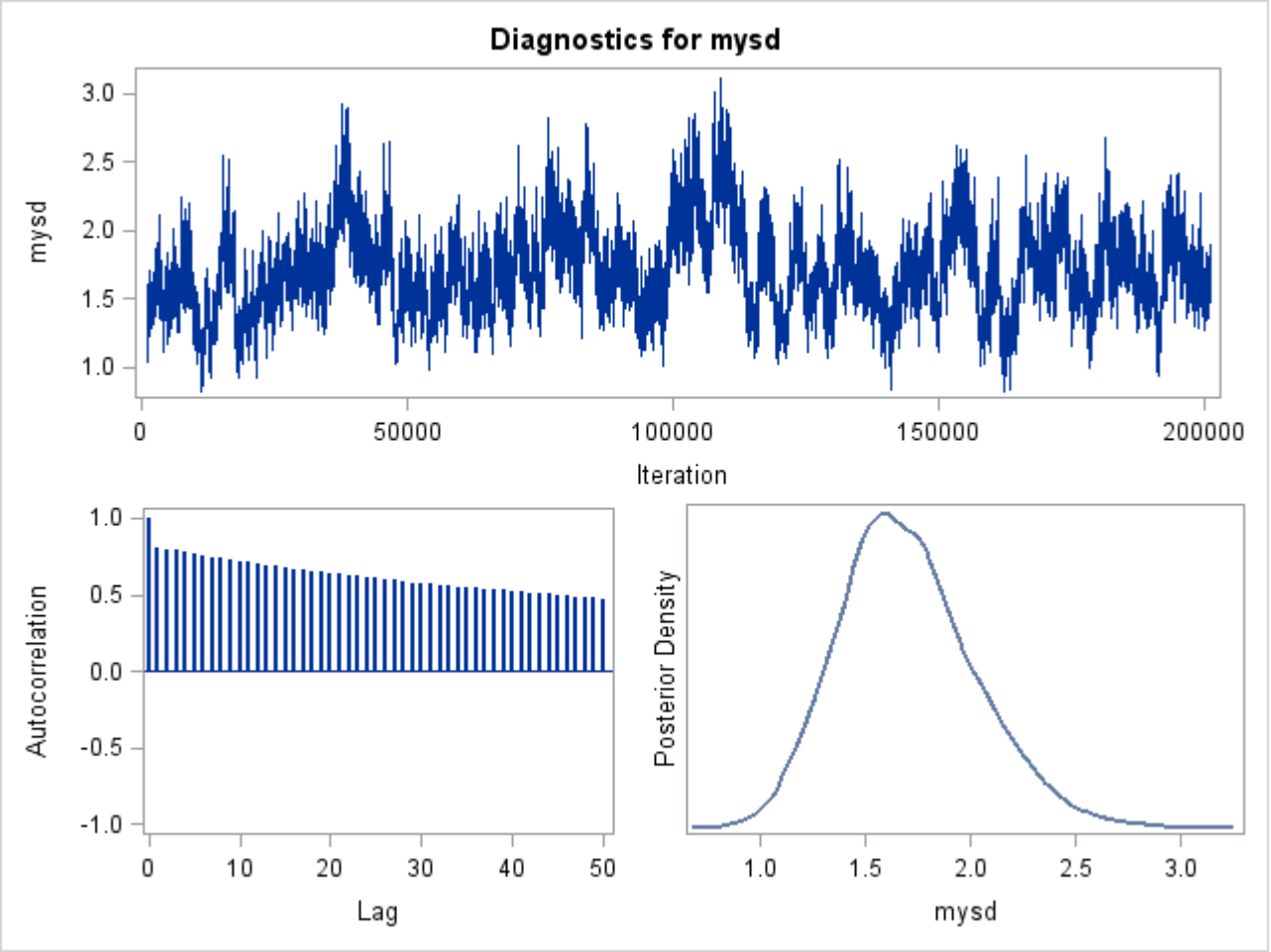
```
proc mcmc data=betablock3  nmc=200000 thin=20 seed=246810 monitor=(mysd);
random Studyeffect ~general(0) subject=Study init=(0) ;
random Treat ~general(0) subject=Trt  init=(0) zero=last monitor=(Treat);
parms logsd 0;
prior logsd ~ general(logsd,lower=log(0.01) upper=log(5));
mysd=exp(logsd);
array zero[3] (0,0,0);
array RE[3];
random RE ~mvnar(zero,sd=mysd/sqrt(2),0) subject=study;
sum=0;
do i=1 to narm;
     sum=sum+RE[i];        ← Take off the average of the random effects
end;
Mu=  Studyeffect + Treat + RE[index] - sum/narm;
P=1-(1/(1+exp(mu)));
model R ~ binomial(n=N, p=P);
run;
```

Omega (2) route

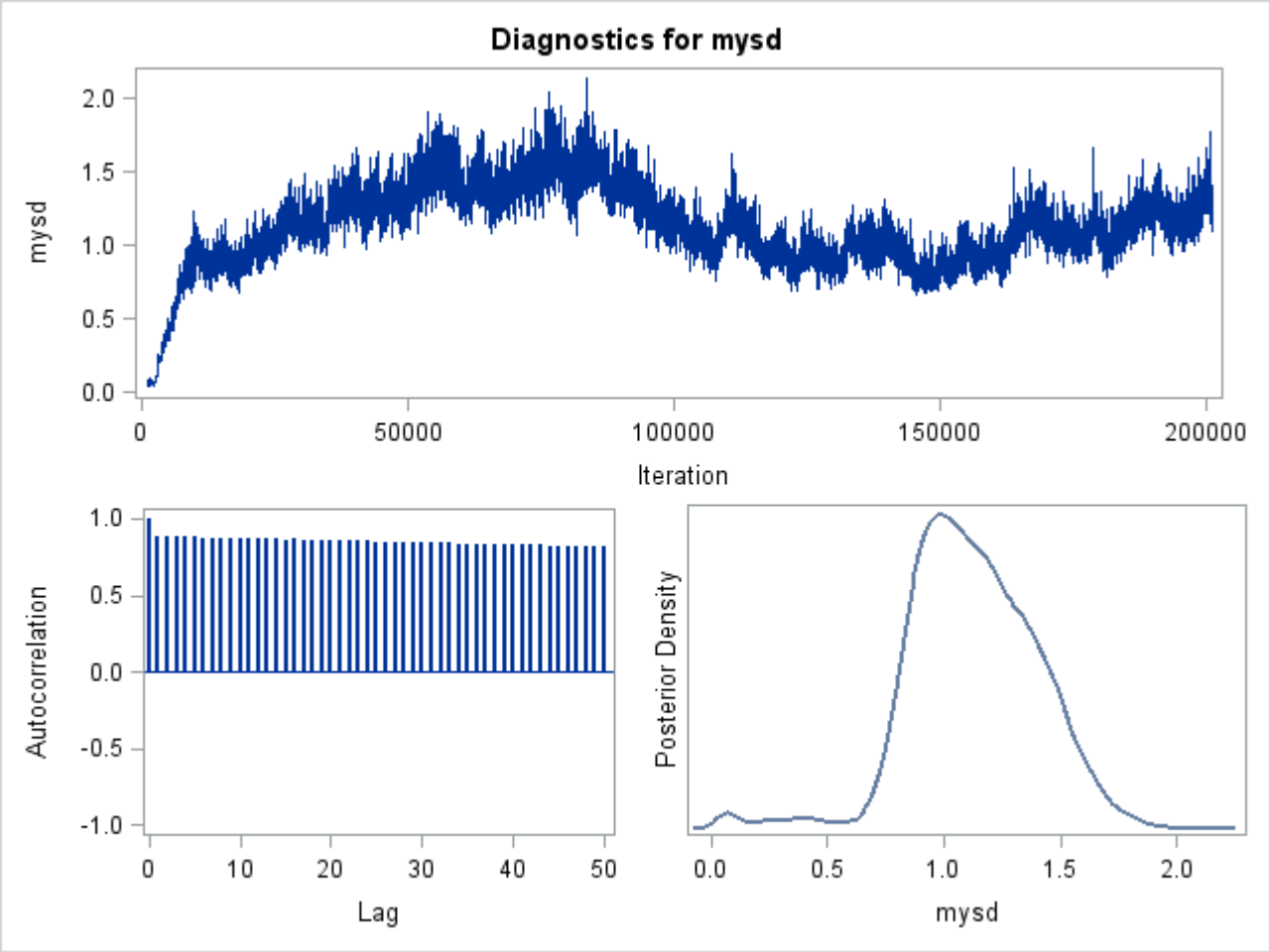
- Neither of these mixes well. (See next two slides)

- But we do not need to go down either of these routes.

Method 1


Diagnostics for mysd
mysd
3.0
2.5
2.0
1.5
1.0
0
50000
100000
150000
200000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
1.0
1.5
2.0
2.5
3.0
mysd

Method 2


Diagnostics for mysd
mysd
2.0
1.5
1.0
0.5
0.0
0
50000
100000
150000
200000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.0
0.5
1.0
1.5
2.0
mysd

MCSE is larger as one might expect.

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
mysd	10000	1.1267	0.2775	0.9492	1.1147	1.3175
Treat_A	10000	-0.7946	0.4027	-1.0716	-0.7788	-0.5130
Treat_B	10000	-0.6554	0.2065	-0.7997	-0.6592	-0.5153

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
mysd	0.0698	0.2775	0.2514
Treat_A	0.0705	0.4027	0.1750
Treat_B	0.0221	0.2065	0.1070

Summary Binomial data (Bayesian)

- MCMC code is very simple.
- Using SAS 9.3, no need to change program from application to application as long as variable names, Study, Trt, R and N remain unchanged.
- Do monitor the diagnostics plots and MCSE.
  - Difficult data sets will make mixing difficult and require long runs with lots of thinning. [See Workshop example.]

Other forms of Data

- Lastly we consider two further types of source data.
- Count data.
- Time to event .

COUNT DATA

Poisson data

- Not common in meta-analyses.
- Often total Count of events and Total length of exposure in each arm, at study level.
- Log link with Log[Exposure] as offset.
- Random effect on Study*Treatment has similar effect as a random effect for over-dispersion.
- PQL with REML (default RSPL) should work well. Better than Gaussian Quadrature with Omega (2).
- Bayesian straightforward adaption to MCMC code.
- But ...

Over-dispersion

- But what about over-dispersion?

- Within a trial over-dispersion usually handled by using estimated scale factor, Negative Binomial distribution, or a normal random effect on linear predictor.

- Usually represents variability in rate from subject to subject (frailty).
- This subtlety is sometimes lost in reporting.

- Say we have total Count in each arm in a study and the relative exposure.
- Use Log link function.
- Use log(exposure) as offset from arm to arm.
- Random effect for frailty should have variance proportional to 1/N the number of subjects in arm.
- Heterogeneity (random effects model) will have similar random effect but with constant variance.
- It is difficult to estimate both from study level data. Better to use assessment of frailty from within each study (as with Normal data).
  - Dispersion parameter from Negative Binomial.
  - Scale parameter if trial uses Poisson with estimated scale.

- Negative Binomial
  - $V(Y) = (\mu + k \mu^2)$ where k is scale parameter.
  - Some confusion as r=1/k is often quoted as Negative Binomial parameter.
  - Use individual k from each trial, or perhaps share across trials.
  - GLIMMIX can handle single scale parameter. Not separate for each trial.
  - Possibly set $SE=\text{sqrt}(\mu + k \mu^2)$ and treat as Normal with MIXED.
  - Or use MCMC and have separate Negative Binomial models with known k for each trial. If SE for k is known include as prior.
- Beware use of simple Poisson regression models.
  - Underestimating within trial variation will increase between study variation. But overall it will inflate Type 2 error for RE model.

- Perhaps simply use estimate of hazard ratios and normal approximation. See next section.

- Should we be worrying about over-dispersion in Binomial data?

TIME TO EVENT DATA

Time to event data.
Usually modelled in terms of the hazard function.

- The underling model is one of proportional hazard

  T is the time of the event.
- f(t) is density function for T.
- F(t) is distribution function for T, $F(t) = \int_0^t f(u)du$.
- S(t) is survivor function, $S(t) = 1 - F(t)$.
- h(t) is hazard function, $h(t) = f(t)/S(t)$.
- H(t) is cumulative hazard function, $H(t) = \int_0^t h(u)du = -log(S(t))$.

  So $F(t) = 1 - exp(-H(t))$

Proportional hazards within trial

The hazard for treatment k in trial i is

$$h_{i,k}(t) = h_i(t)exp(\beta_k)$$

while the cumulative hazard is

$$H_{i,k}(t) = \int_0^t h_i(u)exp(\beta_k)du = H_i(t)exp(\beta_k)$$

where $h_i(t)$ and $H_i(t)$ are the equivalent functions for control arm in trial i.

Then

$$F_{ik}(t) = 1 - exp(-H_i(t)exp(\beta_k))$$

Hazard data

Nearly always presented as comparison to some internal reference (placebo). For each comparison to placebo, have estimate of log(hazard ratio) Y_{ik} and its standard error S_{ik}.

$$log\left(\frac{h_{i,k}(t)}{h_i(t)}\right) = log(h_i(t)) + \beta_k - log(h_i(t)) = \beta_k$$

So assume that $Y_{ik} \sim N(\beta_k, S_{ik}^2)$

But Y_{ik} is not independent of Y_{ij} as both comparisons are made to the same control in the same study.

So need to treat $\mathbf{Y}_i$ as Multivariate normal and include the covariance $Cov(Y_{ik}, Y_{ij})$ which is not usually available in the literature.

Hazard data

Correlation of multiple comparisons to same control within a trial.

For literature reviewed data, one could approximate the *correlation* from the sample sizes.

$$Corr(Y_{ik}, Y_{ij}) \simeq \sqrt{\frac{n_1 n_2}{(n_0 + n_1)(n_0 + n_2)}}$$

which will be 0.5 for equal sized arms.

Easy route

- Build data back to arm level, rather than difference between arms.
- Treat Study effect as fixed with flat prior (important for this to work).
- Set $Y_1=0$ for reference arm in this study.
- Set Y_i= Log Hazard Ratio for treatment i versus reference.
- Set $SE(Y_i)^2=SE(LHR_i)^2\ N_1/(N_1+N_i)$

while set $SE(Y_1)^2$ as any of $SE(LHR_i)^2\ N_i/(N_1+N_i)$ which should all be very similar.

Then proceed as for normal data.

Example: Woods et al, Mortality in COPD.

Table 1 Count Statistics

Author/Trial (Date)	Treatment	r (deaths)	N (patients)
Boyd (1997) [12]	Salmeterol	1	229
	Placebo	1	227
Calverly/TRISTAN (2003) [13]	Fluticasone	4	374
	Salmeterol	3	372
	SFC	2	358
	Placebo	7	361
Celli (2003) [14]	Salmeterol	1	554
	Placebo	2	270

Table 2 Hazard ratio and log hazard ratio statistics

Author/Trial (Date)	Treatment	Base	HR	HR_{LCI}	HR_{UCI}	$\ln(HR)$	$se(\ln(HR))$
Burge/ISOLDE (2000) [15]	Fluticasone	Placebo	0.76	0.51	1.13	-0.276	0.203
Calverly/TORCH (2007) [16]	SFC	Placebo	0.811	0.670	0.982	-0.209	0.098
	Salmeterol	Placebo	0.857	0.710	1.035	-0.154	0.096
	Fluticasone	Placebo	1.056	0.883	1.264	0.055	0.092
	SFC	Salmeterol	0.946	0.777	1.151	-0.056	0.100
	SFC	Fluticasone	0.768	0.636	0.927	-0.264	0.096

Woods, Hawkins & Scott (2010) BMC Med. Res. Methodology, 10:54.

Part of original data from Baker et al (2009) Pharmacotherapy, 29(8)891-905.

Combining Time to event with event rate data.

Frequency data

Here the data are the count of events R_{ik}/N_{ik} in arm k for trial i , including the control arm 0.

$$F_{ik}(t) = 1 - exp(-H_i(t)exp(\beta_k)) = 1 - exp(-exp(\alpha_i + \beta_k))$$

where α_i is the log(cumulative hazard) at termination of trial i in the control arm.

So we model the observed data as

$$R_{ik} \sim Bin(exp(-exp(\alpha_i + \beta_k)), N_{ik})$$

Frequency data

- Use Complementary log-log link with Binomial error.
- Note that this is close to logistic when rate is small.
- But we are assuming a different model from the usual log-odds-ratio model.

Computation.
Different types of data.

- In GLIMMIX can we specify distribution and link function in the data.

- In MCMC we can calculate the log-likelihood directly for the different types of data and specify using,

Model y ~ general(log likelihood);

- In Winbugs, separate arrays for each type of data, with associated distribution declarations.

```
data LHR;
INPUT Study Trt  LHR
SELHR ;
WT=1/(SELHR**2);
Dist=1;
Link=1;
N=1;
Y=LHR;
datalines;
1 1  0   0.066
1 2  0.055   0.063
1 3  -0.154   0.070
1 4  -0.209   0.072
2 1  0   0.1435427
2 2 -0.276   0.1435427
;
run;
```

```
data Binary;
input Study TRT R N;
WT=1;
Dist=3;
Link=5;
Y=R;
datalines;
3 3  1 229
3 1  1 227
4 2  4 374
4 3  3 372
4 4  2 358
4 1  7 361
5 3  1 554
5 1  2 270
;
run;
```

GLIMMIX (Fixed model)

```
proc glimmix data=alldata;

class Trt Study ;

model Y/N = Trt study /Dist=BYOBS(Dist) Link=BYOBS(Link)  Solution
     ddfm=None;

parms 1 / Hold=1;

weight WT;

estimate "SFC - Placebo"  Trt -1 1 0 0  /CL;

estimate "Sal - Placebo"  Trt -1 0 1 0  /CL;

estimate "FP  - Placebo"  Trt -1 0 0 1  /CL;

ods output estimates=est;

run;
```

Transform back to Hazard Ratios

```
* Transform back onto the HR scale;
data est2;
set est;
HR=exp(ESTIMATE);
LHR=exp(lower);
UHR=exp(upper);
run;
```

Glimmix
Hazard ratio estimates

Label	HR	LowerHR	UpperHR
SFC - Placebo	0.77214	0.64125	0.92974
Sal - Placebo	0.81568	0.67992	0.97854
FP - Placebo	0.98552	0.83839	1.15848

Table 4 Network meta-analysis results

Comparator	Hazard ratio (95% CI) vs. placebo - fixed effects
Fluticasone	0.99 (0.84, 1.16)
Salmeterol	0.82 (0.68, 0.98)
SFC	0.78 (0.64, 0.93)
random effect SD	-
DIC	25.25

MCMC fixed effects

```
proc mcmc data=Alldata  NBI=10000 NMC=100000 thin=10 Stats=all
     mssing=AC seed=12345 monitor=(SFC_Plac Sal_Plac FP_Plac);

random Study_eff ~ general(0) subject=study init=(0);

random trt_eff ~general(0) subject=trt zero=first monitor=(trt_eff) init=(0);

mu= Study_eff + Trt_eff ;

if R>= 0 then do;

     * Binomial data;

     p = logistic(mu);

     ll= logpdf("Binomial",r,p,n);

end;

else do;

     * Hazard ratio data;

     ll= logpdf("Normal",lhr,mu,selhr);

end;

model study ~ general(ll);
```

```
array effect[4];
effect[trt]=trt_eff;
beginnodata;
    SFC_Plac=exp(effect[4]-effect[1]);
    Sal_Plac=exp(effect[3]-effect[1]);
    FP_Plac=exp(effect[2]-effect[1]);
endnodata;
run;
```

Side effect of using RANDOM statement for treatment effect.

- Need to copy estimates into an array and then use these to build the hazard ratios.

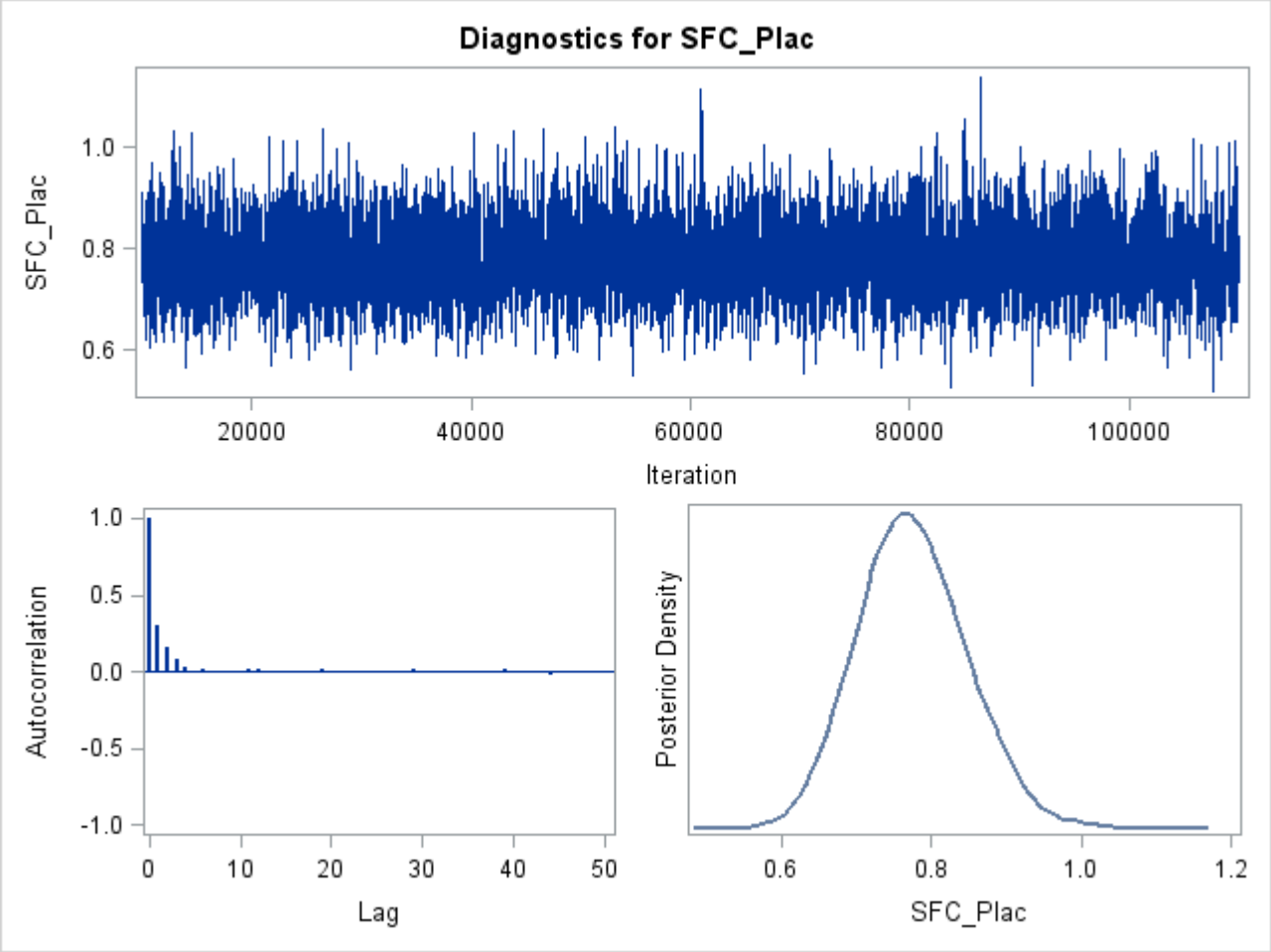
Fixed effects using MCMC

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
SFC_Plac	10000	0.7757	0.0731	0.7256	0.7725	0.8233
Sal_Plac	10000	0.8202	0.0764	0.7673	0.8161	0.8701
FP_Plac	10000	0.9895	0.0814	0.9319	0.9864	1.0432
trt_eff_2	10000	-0.0139	0.0822	-0.0705	-0.0137	0.0423
trt_eff_3	10000	-0.2026	0.0929	-0.2648	-0.2032	-0.1391
trt_eff_4	10000	-0.2584	0.0942	-0.3208	-0.2581	-0.1944

Table 4 Network meta-analysis results

Comparator	Hazard ratio (95% CI) vs. placebo - fixed effects
Fluticasone	0.99 (0.84, 1.16)
Salmeterol	0.82 (0.68, 0.98)
SFC	0.78 (0.64, 0.93)
random effect SD	-
DIC	25.25

Diagnostics are beautiful.



Random effects GLIMMIX

```
proc glimmix data=alldata method=RSPL;;

class Trt Study ;

model Y/N = Trt study /Dist=BYOBS(Dist) Link=BYOBS(Link)  Solution
     ddfm=KR;

random intercept /subject=Study*Trt ;

parms 1 1 / Hold=(2);

weight WT;

estimate "SFC - Placebo"  Trt -1 0 0 1  /CL ;

estimate "Sal - Placebo"  Trt -1 0 1 0  /CL ;

estimate "FP  - Placebo"  Trt -1 1 0 0  /CL ;

ods output estimates=estr;

run;
```

Without and with DDFM=KR.

Label	HR	LowerHR	UpperHR
SFC - Placebo	0.71404	0.45656	1.11673
Sal - Placebo	0.75436	0.48689	1.16874
FP - Placebo	0.90412	0.62200	1.31419

Label	HR	LowerHR	UpperHR
SFC - Placebo	0.71404	0.04978	10.2415
Sal - Placebo	0.75436	0.04901	11.6119
FP - Placebo	0.90412	0.15717	5.2010

Comparator	Hazard ratio (95% CI) vs. placebo - fixed effects	Hazard ratio (95% CI) vs. placebo - random effects
Fluticasone	0.99 (0.84, 1.16)	0.89 (0.39, 1.42)
Salmeterol	0.82 (0.68, 0.98)	0.73 (0.29, 1.23)
SFC	0.78 (0.64, 0.93)	0.69 (0.26, 1.21)
random effect SD	-	0.36 (0.31)
DIC	25.25	25.73

Always beware KR d.f. That are this small.

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
SFC - Placebo	-0.3368	0.2096	1	-1.61	0.3544	0.05	-3.0001	2.3265
Sal - Placebo	-0.2819	0.2152	1	-1.31	0.4150	0.05	-3.0158	2.4520
FP - Placebo	-0.1008	0.1826	1.138	-0.55	0.6697	0.05	-1.8504	1.6488

MCMC [Random effect using SAS 9.3]

```
parms logsd 0;
prior logsd ~ general(logsd, upper=log(5));
mysd=exp(logsd);
random randeff ~ normal(0, sd=mysd/sqrt(2)) subject=_obs_ ;

mu= Study_eff + Trt_eff + randeff;
```

Hazard ratios from random effects model

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
SFC_Plac	10000	0.7027	0.2973	0.5372	0.6871	0.8101
Sal_Plac	10000	0.7143	0.2394	0.5672	0.7159	0.8479
FP_Plac	10000	0.8946	0.2659	0.7374	0.8912	1.0259
mysd	10000	0.3918	0.3206	0.1766	0.3115	0.5134

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
SFC_Plac	0.050	0.2782	1.4113	0.2146	1.2378
Sal_Plac	0.050	0.2525	1.2214	0.1957	1.1397
FP_Plac	0.050	0.4236	1.4752	0.3631	1.3769
mysd	0.050	0.0309	1.2449	0.00793	1.0135

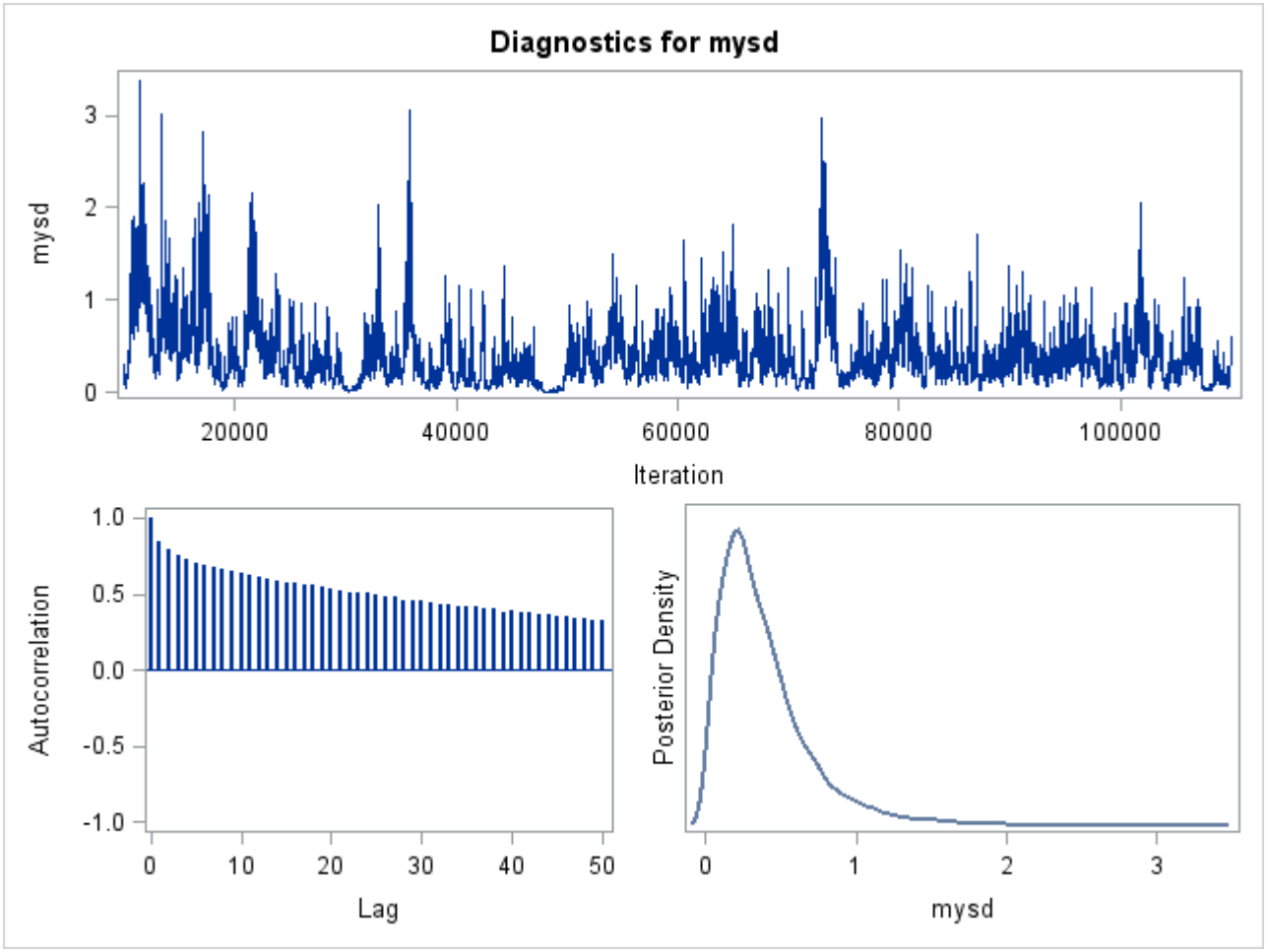
Comparator	Hazard ratio (95% CI) vs. placebo - fixed effects	Hazard ratio (95% CI) vs. placebo - random effects
Fluticasone	0.99 (0.84, 1.16)	0.89 (0.39, 1.42)
Salmeterol	0.82 (0.68, 0.98)	0.73 (0.29, 1.23)
SFC	0.78 (0.64, 0.93)	0.69 (0.26, 1.21)
random effect SD	-	0.36 (0.31)
DIC	25.25	25.73

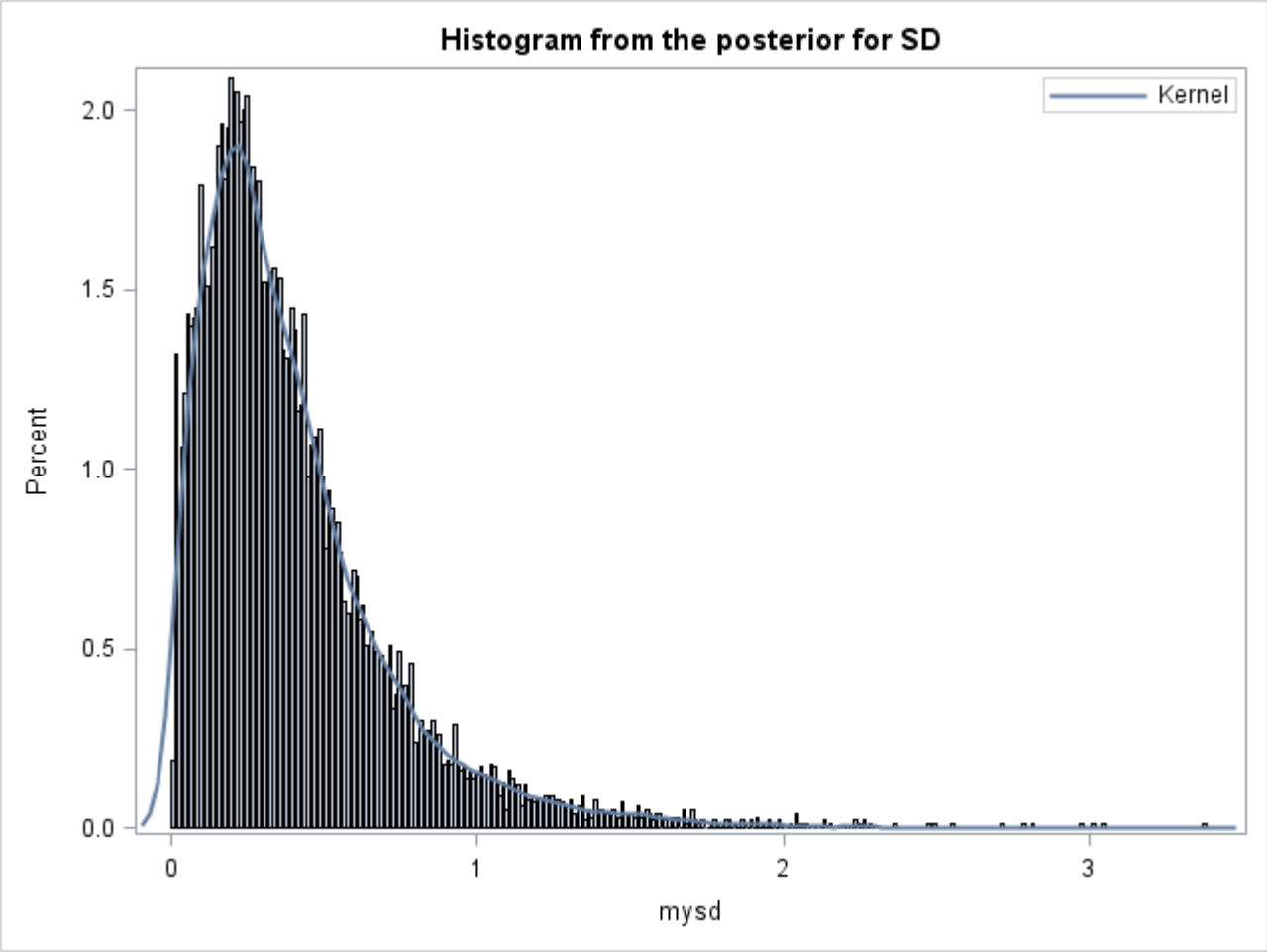
Note that consistent within MCSE.

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
SFC_Plac	0.0224	0.2973	0.0755
Sal_Plac	0.0183	0.2394	0.0766
FP_Plac	0.0194	0.2659	0.0729
mysd	0.0288	0.3206	0.0898

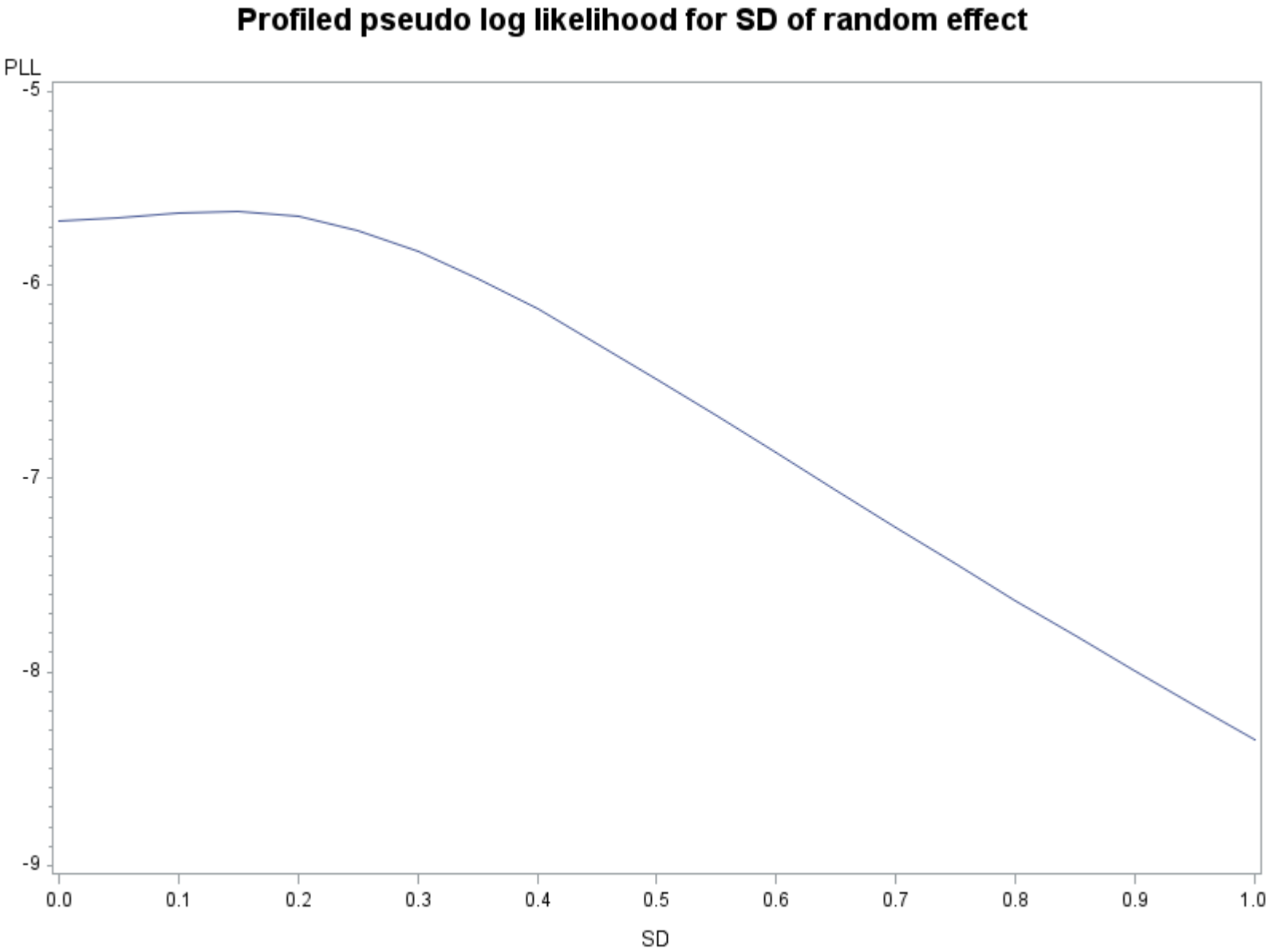
Only took 5 seconds so can afford to throw lots of iterations at it.

Mixing was poor




Histogram from the posterior for SD
Kernel
Percent
2.0
1.5
1.0
0.5
0.0
0
1
2
3
mysd

Why it is difficult!

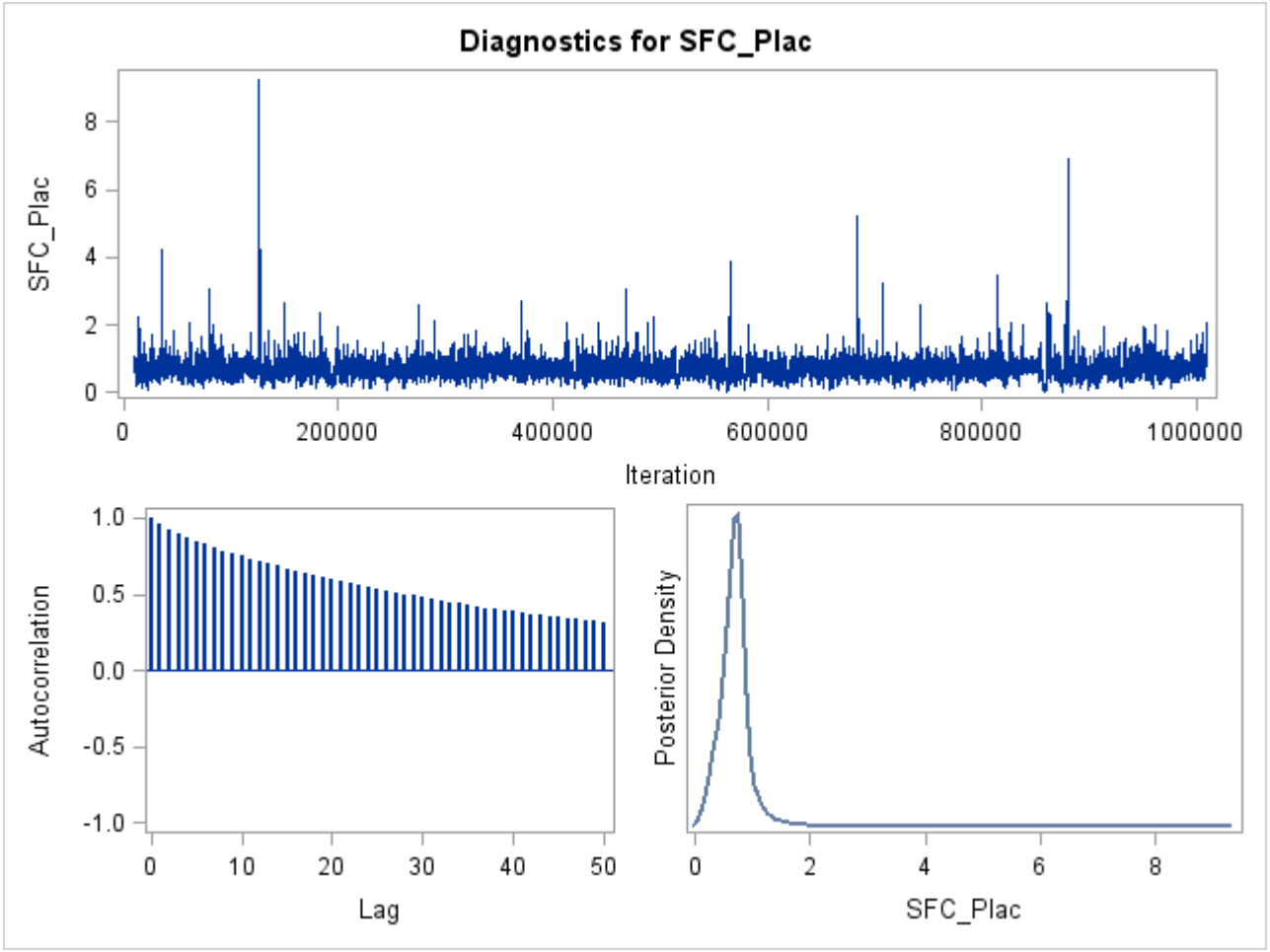


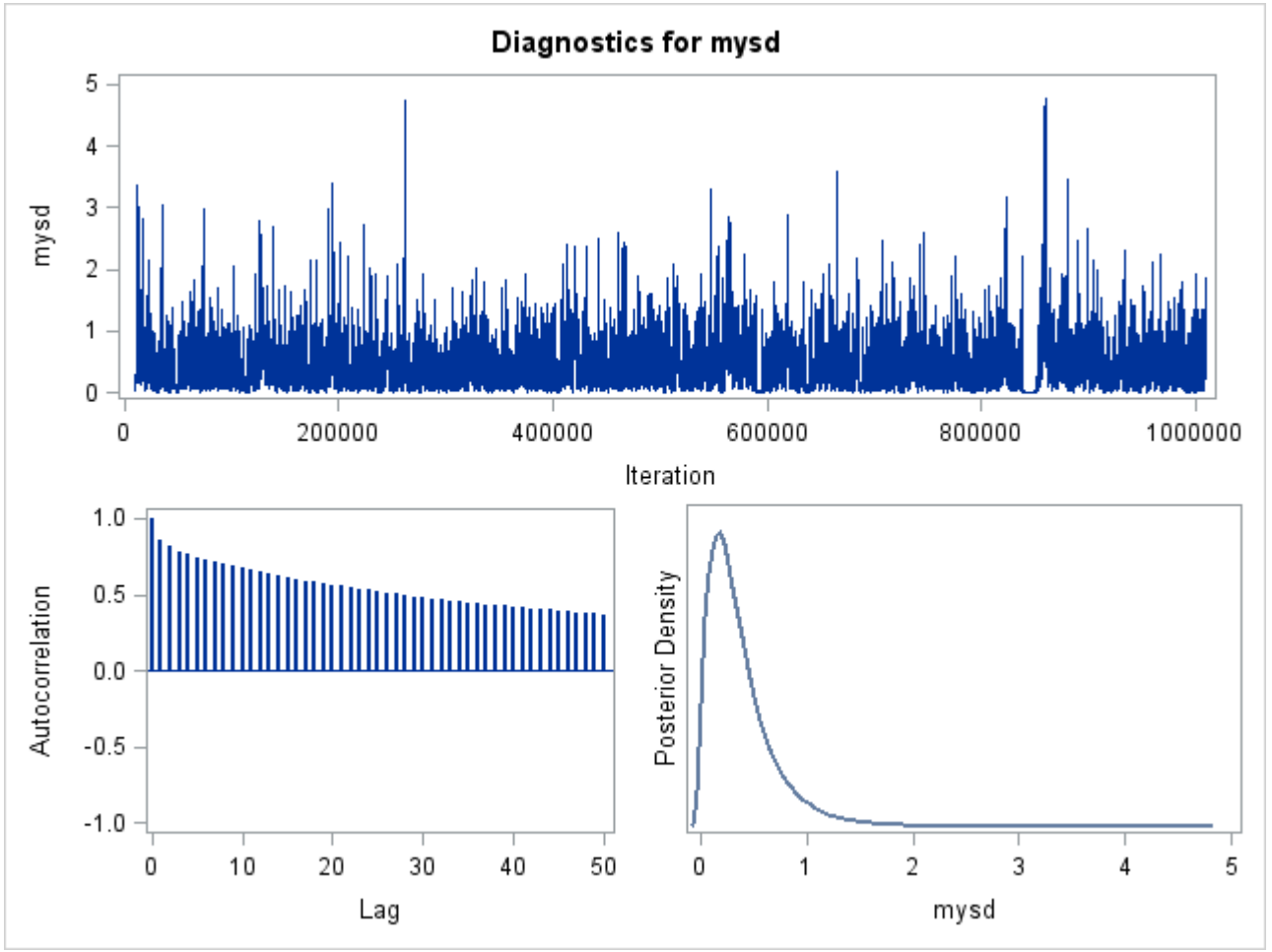
Using NBI=10,000 NMC=1,000,000 thin=10

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
SFC_Plac	100000	0.6968	0.3091	0.5446	0.6893	0.8083
Sal_Plac	100000	0.7292	0.2462	0.5884	0.7306	0.8508
FP_Plac	100000	0.8942	0.2988	0.7433	0.8917	1.0178
mysd	100000	0.3772	0.3241	0.1563	0.2944	0.5025

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
SFC_Plac	0.050	0.2287	1.2886	0.1418	1.1285
Sal_Plac	0.050	0.2712	1.2538	0.2151	1.1586
FP_Plac	0.050	0.3838	1.4696	0.3250	1.3744
mysd	0.050	0.0200	1.2209	0.00199	0.9997

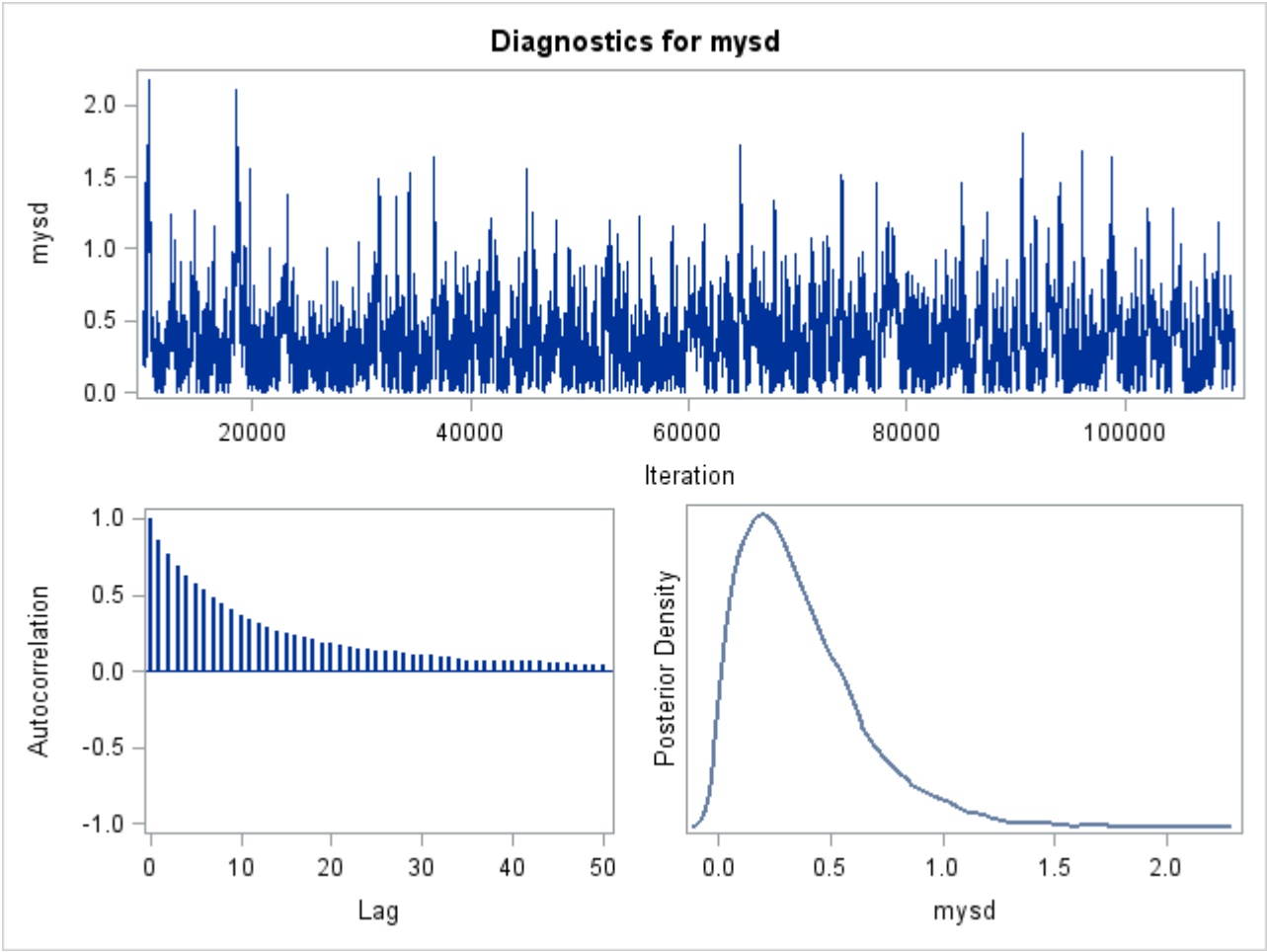
Comparator	Hazard ratio (95% CI) vs. placebo - fixed effects	Hazard ratio (95% CI) vs. placebo - random effects
Fluticasone	0.99 (0.84, 1.16)	0.89 (0.39, 1.42)
Salmeterol	0.82 (0.68, 0.98)	0.73 (0.29, 1.23)
SFC	0.78 (0.64, 0.93)	0.69 (0.26, 1.21)
random effect SD	-	0.36 (0.31)
DIC	25.25	25.73


Diagnostics for SFC_Plac
SFC_Plac
8
6
4
2
0
0
200000
400000
600000
800000
1000000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
2
4
6
8
SFC_Plac


Diagnostics for mysd
mysd
5
4
3
2
1
0
0
200000
400000
600000
800000
1000000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0
1
2
3
4
5
mysd

Final trick to help mix

```
parms logsd 0 /slice;

prior logsd ~ general(logsd, upper=log(5)) ;

mysd=exp(logsd);

random randeff ~ normal(0, sd=1) subject=_obs_ ;


mu= Study_eff + Trt_eff + randeff*mysd/sqrt(2);
```


Diagnostics for mysd
mysd
2.0
1.5
1.0
0.5
0.0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.0
0.5
1.0
1.5
2.0
mysd

Improvement in effective sample size

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
SFC_Plac	233.0	42.9210	0.0233
Sal_Plac	233.6	42.8141	0.0234
FP_Plac	299.9	33.3411	0.0300
mysd	435.9	22.9411	0.0436

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
SFC_Plac	175.6	56.9574	0.0176
Sal_Plac	170.3	58.7182	0.0170
FP_Plac	188.2	53.1217	0.0188
mysd	124.1	80.5837	0.0124

Summary using SAS for indirect comparisons.

- As long as we express model as a GLMM then we can use either GLIMMIX or MCMC procedures.
- GLIMMIX
  - Preferably use PQL with REML (method=RSPL) but be aware of the problem with Binary data and also with very small rates for binomial data.
- MCMC
  - Issue of aliasing of fixed and random effects handled automatically as long as use flat priors for treatment and study.
  - The Random statement in SAS 9.3 helps makes code transparent.

Summary

- Random effects model have fixed margins for Study and Treatment but Random interaction.
  - This is the source of much of the misunderstandings.
- Use Winbugs if that makes life easy for you.
- But GLIMMIX and MCMC procedures make it easy in SAS.
- GLIMMIX with RSPL will often give good quick answers without having to mess around with MCMC.

Some final thoughts. (1)

- Broken networks
  - Do not fix them by simply merging trials (Bad Programmer's solution).
  - Do not fix by using random study effects (Bad Statistician's solution)
  - You cannot bridge.

Some final thoughts. (2)

- What makes MCMC and Winbugs difficult to control.
  - A flat likelihood for the random effect SD near zero.
  - A network with a weak bridge.

- What makes GLIMMIX with PQL behave less well.
  - Binomial data with very low frequencies.

WORKSHOP 3

Workshop

- Binary data / Binomial data.

- Here we will experiment with a more complex network.

Cipriani et al

Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al.

Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis.

Lancet 2009 Feb 28;**373(9665):746-758;**

Cipriani et al

Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al.

Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis.

Lancet 2009 Feb 28;**373(9665):746-758;**

The acceptability data are used as an example data set in

Interpreting Indirect Treatment Comparisons and Network Meta-Analysis forHealth-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1.

Value in Helath 14 (2011) 417– 428; doi:10.1016/j.jval.2011.04.002

Summary

“Background Conventional meta-analyses have shown inconsistent results for efficacy of second-generation antidepressants.

We therefore did a multiple-treatments meta-analysis, which accounts for both direct and indirect comparisons, to assess the effects of 12 new-generation antidepressants on major depression.”

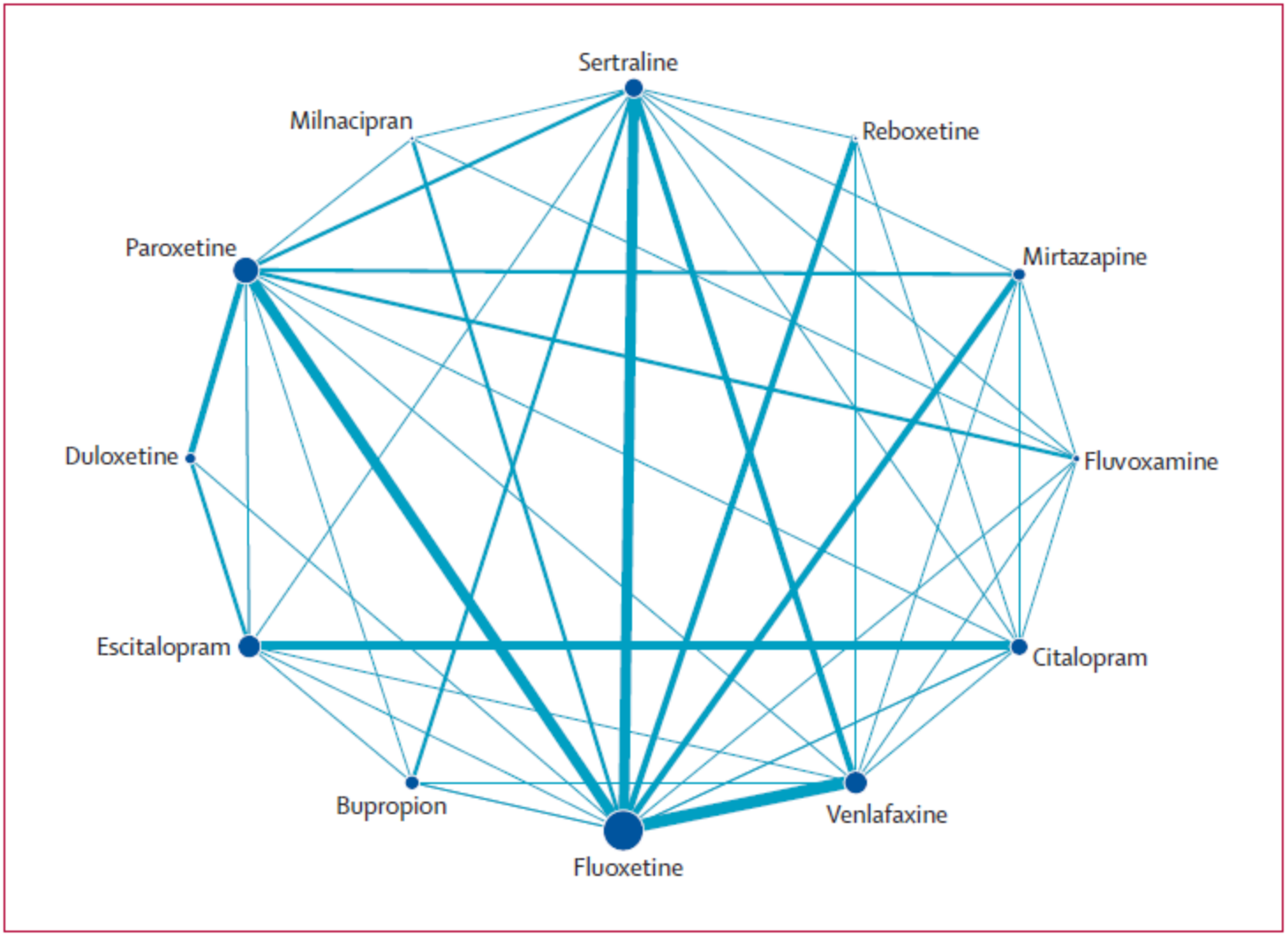


Figure 2: **Network of eligible comparisons for the multiple-treatment meta-analysis for efficacy (response rate)** The width of the lines is proportional to the number of trials comparing each pair of treatments, and the size of each node is proportional to the number of randomised participants (sample size). The network of eligible comparisons for acceptability (dropout rate) analysis is similar.

Studies with Acceptability data.

- 112 studies;
- 226 records (Study*Treat);
- 12 treatments;

Actions

- Follow the steps in the handout.
- Program file is Workshop3.sas
- We will discuss our results at the end.

Random effects model

```
Title1 "Random effect with RSPL (proc GLIMMIX)";

proc glimmix data=Cip2 method=RSPL;

class study Treatment ;

model acceptR/DropN = Study Treatment /link=logit
    dist=bin ddfm=kr;

random intercept /subject=Study*Treatment ;

lsmeans Treatment /diff=control('fluoxetine') cl oddsratios;

run;
```

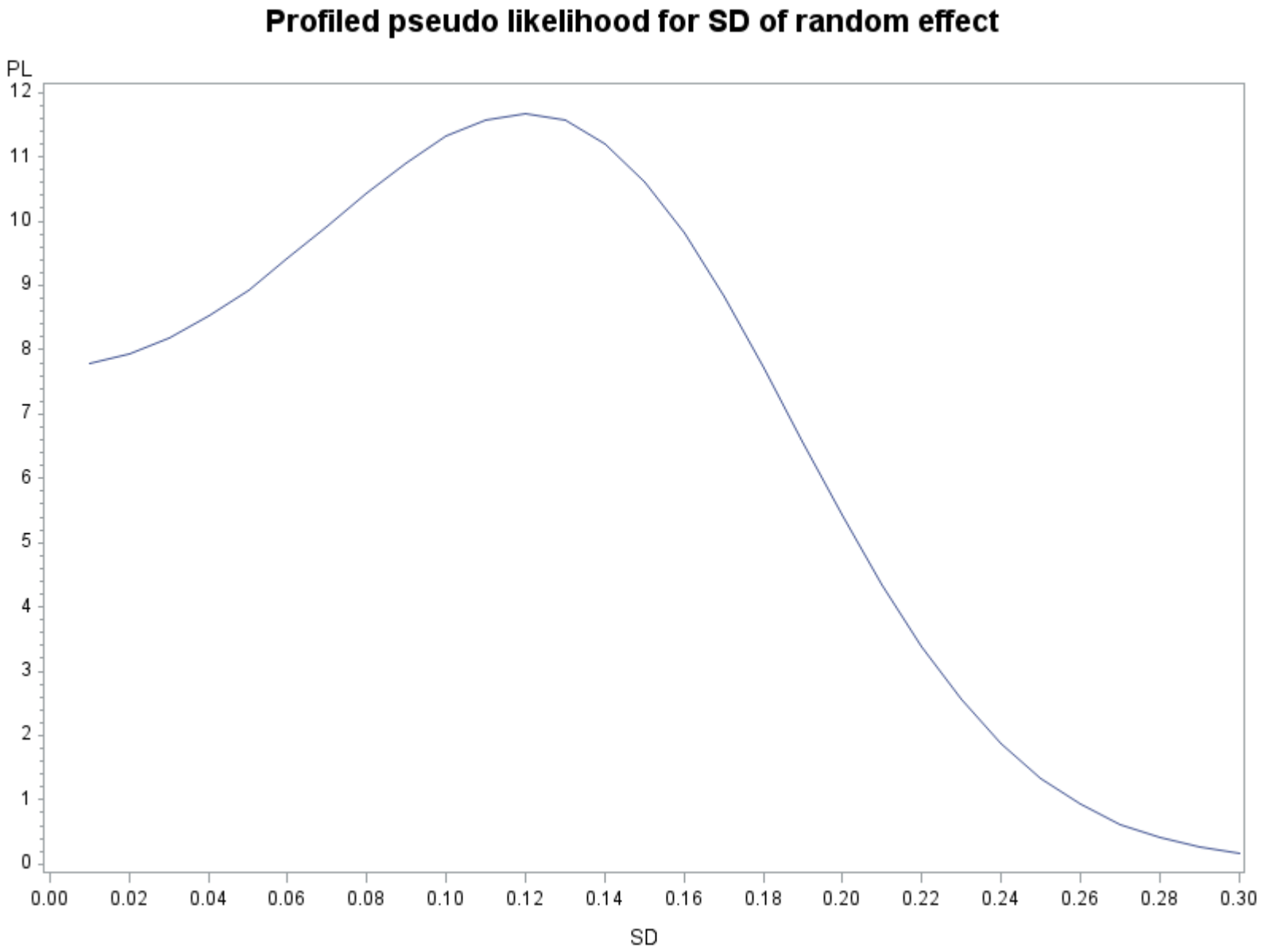
Random effects

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	Study*treatment	0.008043	0.009374

sqrt(0.008043*2) = 0.1268306

Random effects (GLIMMIX)

Differences of treatment Least Squares Means												
treatment	_treatment	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
bupropion	fluoxetine	0.1101	0.09692	43.95	1.14	0.2621	0.05	-0.08524	0.3054	1.116	0.918	1.357
citalopram	fluoxetine	0.1080	0.1010	51.84	1.07	0.2899	0.05	-0.09471	0.3108	1.114	0.910	1.365
duloxetine	fluoxetine	-0.1786	0.1375	44.94	-1.30	0.2005	0.05	-0.4555	0.09829	0.836	0.634	1.103
escitalopram	fluoxetine	0.1716	0.09308	53.95	1.84	0.0708	0.05	-0.01505	0.3582	1.187	0.985	1.431
fluvoxamine	fluoxetine	-0.2033	0.1380	103	-1.47	0.1437	0.05	-0.4770	0.07037	0.816	0.621	1.073
milnacipran	fluoxetine	-0.03215	0.1562	69.31	-0.21	0.8375	0.05	-0.3437	0.2794	0.968	0.709	1.322
mirtazapine	fluoxetine	-0.03296	0.1136	62.62	-0.29	0.7726	0.05	-0.2600	0.1941	0.968	0.771	1.214
paroxetine	fluoxetine	-0.09970	0.07216	39.02	-1.38	0.1750	0.05	-0.2457	0.04627	0.905	0.782	1.047
reboxetine	fluoxetine	-0.3582	0.1415	55.91	-2.53	0.0142	0.05	-0.6417	-0.07474	0.699	0.526	0.928
sertraline	fluoxetine	0.1226	0.08890	77.63	1.38	0.1719	0.05	-0.05441	0.2996	1.130	0.947	1.349
venlafaxine	fluoxetine	-0.06394	0.07687	62.92	-0.83	0.4087	0.05	-0.2176	0.08967	0.938	0.804	1.094


Profiled pseudo likelihood for SD of random effect
PL
12
11
10
9
8
7
6
5
4
3
2
1
0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
SD

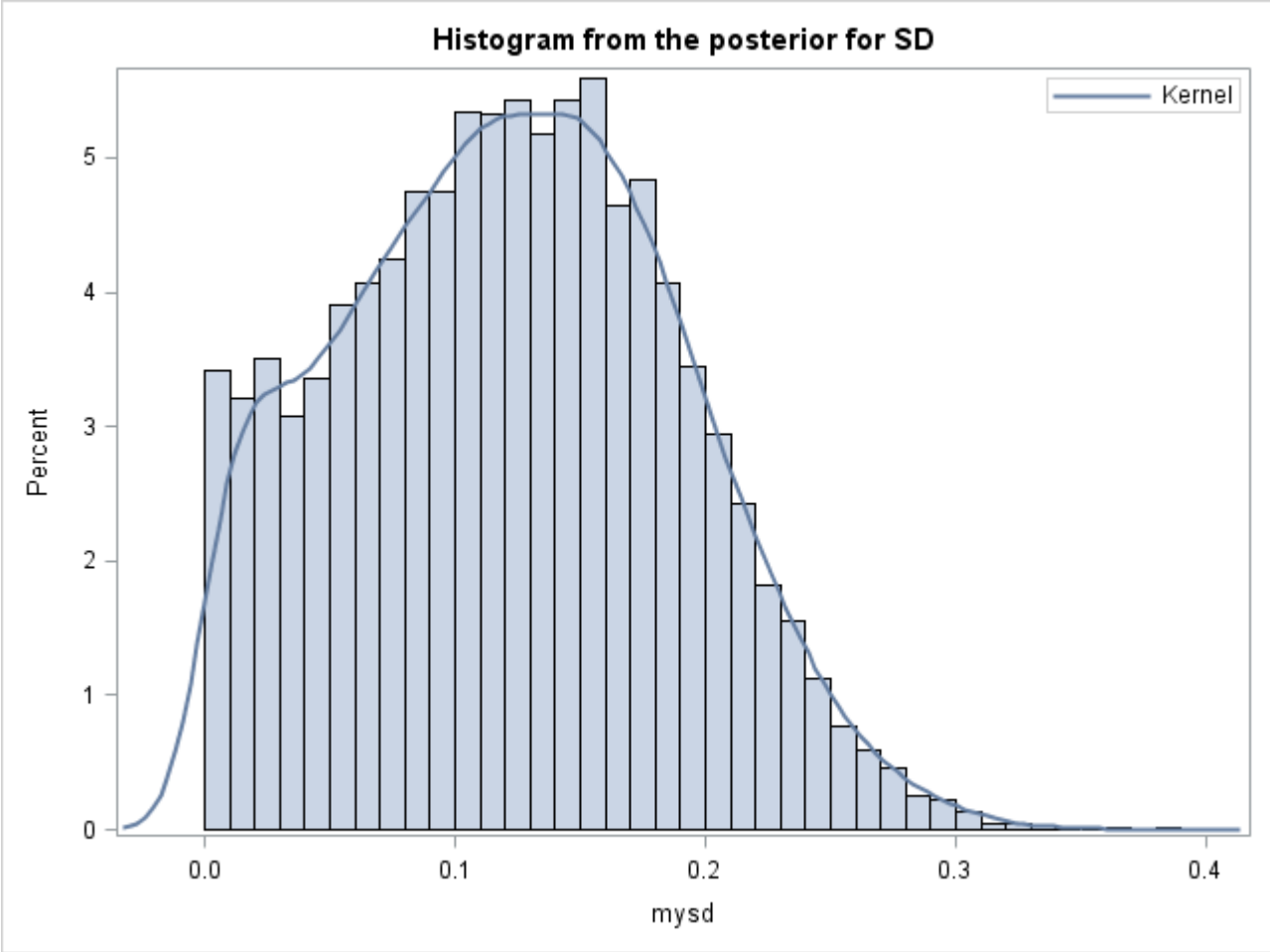
- Now move to MCMC using SAS 9.3

The run below with 200,000 takes 6 minutes. Version in program code uses 20,000 (45secs)

```
proc mcmc data=Cip2  nmc=200000 thin=20 seed=246810 monitor=(mysd OR) outpost=outp1;
random Studyeffect ~general(0) subject=ShortStudy init=(0) ;
random Treat_eff ~general(0) subject=Treatment  init=(0) zero=last; * monitor=(Treat_eff);
parms logsd 0 /slice;
prior logsd ~ general(logsd, upper=log(5));
mysd=exp(logsd);
random RE ~normal(0,sd=1) subject=_OBS_ init=(0);
Mu=  Studyeffect + Treat_eff + RE*mysd/sqrt(2);
P=1-(1/(1+exp(mu)));
model AcceptR ~ binomial(n=DropN, p=P);
array effect[12];
array OR[12];
effect[trt]=treat_eff;
beginnodata;
      do i=1 to 12;
            * contrasts to fluoxetine;
            OR[i]=exp(effect[i]-effect[5]);
      end;
endnodata;
run;
```

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
mysd	10000	0.1219	0.0662	0.0710	0.1217	0.1697
OR1	10000	1.1221	0.1126	1.0432	1.1163	1.1929
OR2	10000	1.1189	0.1159	1.0379	1.1136	1.1937
OR3	10000	0.8424	0.1212	0.7579	0.8333	0.9170
OR4	10000	1.1865	0.1130	1.1067	1.1816	1.2590
OR5	10000	1.0000	0	1.0000	1.0000	1.0000
OR6	10000	0.8225	0.1161	0.7427	0.8126	0.8935
OR7	10000	0.9784	0.1559	0.8675	0.9665	1.0767
OR8	10000	0.9733	0.1125	0.8950	0.9668	1.0441
OR9	10000	0.9066	0.0663	0.8606	0.9038	0.9492
OR10	10000	0.7030	0.1018	0.6310	0.6955	0.7666
OR11	10000	1.1369	0.1054	1.0642	1.1309	1.2029
OR12	10000	0.9394	0.0734	0.8892	0.9363	0.9861

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
mysd	0.050	0.00707	0.2511	0.000183	0.2319
OR1	0.050	0.9206	1.3660	0.9115	1.3469
OR2	0.050	0.9072	1.3596	0.9101	1.3619
OR3	0.050	0.6292	1.1000	0.6137	1.0771
OR4	0.050	0.9776	1.4248	0.9625	1.4046
OR5	0.050	1.0000	1.0000	1.0000	1.0000
OR6	0.050	0.6207	1.0735	0.6054	1.0538
OR7	0.050	0.7079	1.3115	0.6948	1.2922
OR8	0.050	0.7736	1.2138	0.7605	1.1964
OR9	0.050	0.7836	1.0457	0.7741	1.0337
OR10	0.050	0.5247	0.9198	0.5153	0.9046
OR11	0.050	0.9491	1.3606	0.9404	1.3473
OR12	0.050	0.8052	1.0930	0.7955	1.0809


Histogram from the posterior for SD
Kernel
Percent
0
1
2
3
4
5
0.0
0.1
0.2
0.3
0.4
mysd

Code SAS 9.2

- Need to add index for Study as well as Treatment.
- The random effect RE has lots of levels.
  - Break it's PARMS statement up into several to reduce dimensionality of Metropolis Hastings.
- The SAS 9.3 solution is much faster.

Further Network Meta-Analysis Models

Neil Hawkins, PhD, Cstat
(http://neilhawkins.wordpress.com/)
Vice President, Health Economics, Icon PLC
Honorary Professor, (Health Economics and Health Technology Assessment), University of Glasgow

Case Studies

1. Co-variable adjustment
2. Inconsistency
3. Complex interventions
4. Multiple outcomes
5. Multiple follow-up times
6. Propensity score methods

ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 857–864

Journal of Clinical Epidemiology

A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered

Georgia Salanti[a,b,*], Valeria Marinho[c], Julian P.T. Higgins[a]

[a]*MRC Biostatistics Unit, Cambridge, UK*

[b]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*

[c]*Clinical and Diagnostic Oral Sciences, Barts and The London School of Medicine and Dentistry Queen Mary, University of London, London, UK*

Accepted 3 October 2008

Abstract

Objective: To illustrate the potential and challenges of the simultaneous analysis of a network of trials, using as a case study the investigation of the relative effectiveness of four topical fluoride treatments and two control interventions (placebo and no treatment) in preventing dental caries in children.

Study Design and Setting: We performed multiple-treatments meta-analysis within a Bayesian framework by synthesizing six Cochrane reviews. We explored the compatibility between direct and indirect evidence and adjusted the results using a meta-regression model to take into account differences in the year of randomization across studies.

Results: The validity of our conclusions for the superiority of fluoride toothpaste as indicated from the initial network analysis using Bayesian methods was challenged when we adjusted for possible confounders. The network was dominated by studies comparing placebo with toothpaste, which were older and had been carried out in populations with higher baseline risk than studies involving other fluoride modalities.

Conclusion: After adjusting for possible differences across studies, we did not find clear evidence that any topical fluoride modality is more effective than any other. Multiple-treatments meta-analysis methods allow for more detailed investigations than naïve methods in the analysis of indirect evidence on treatment effects. © 2009 Elsevier Inc. All rights reserved.

Keywords: Meta-analysis; Topical fluoride therapy; Mixed treatment comparisons; Multiple-treatments meta-analysis; Incoherence; Meta-regression; Dental caries

Network of comparisons of fluoride therapies

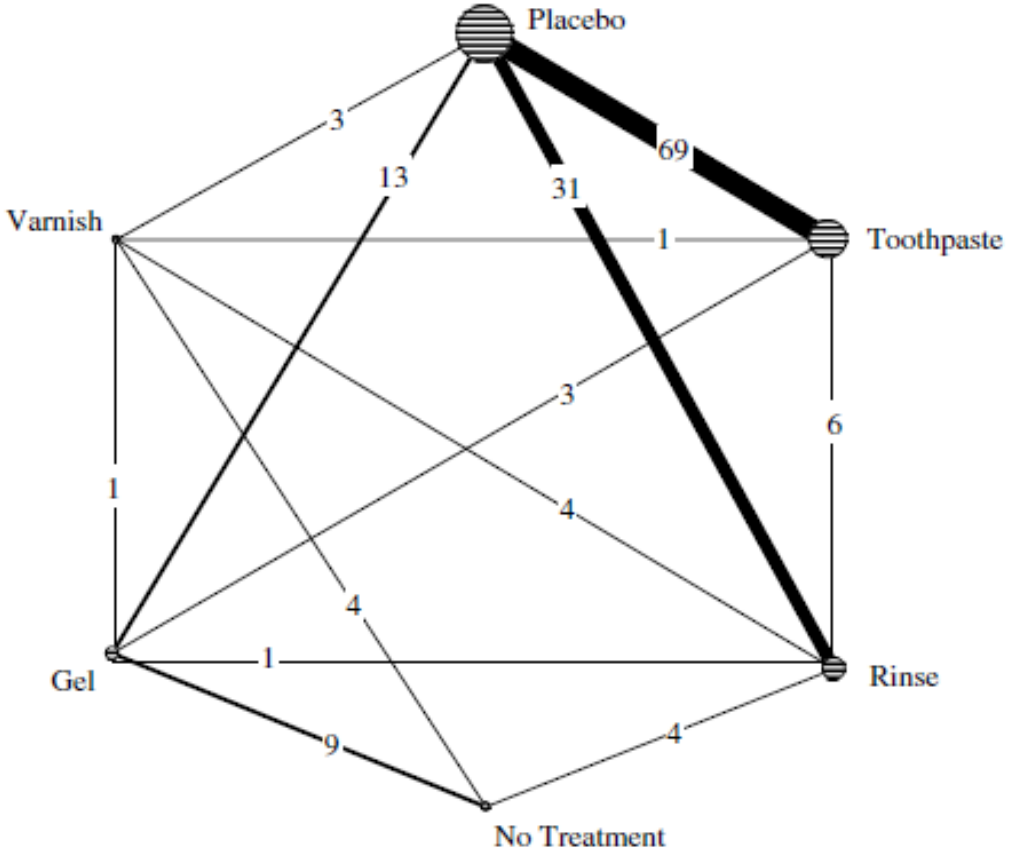


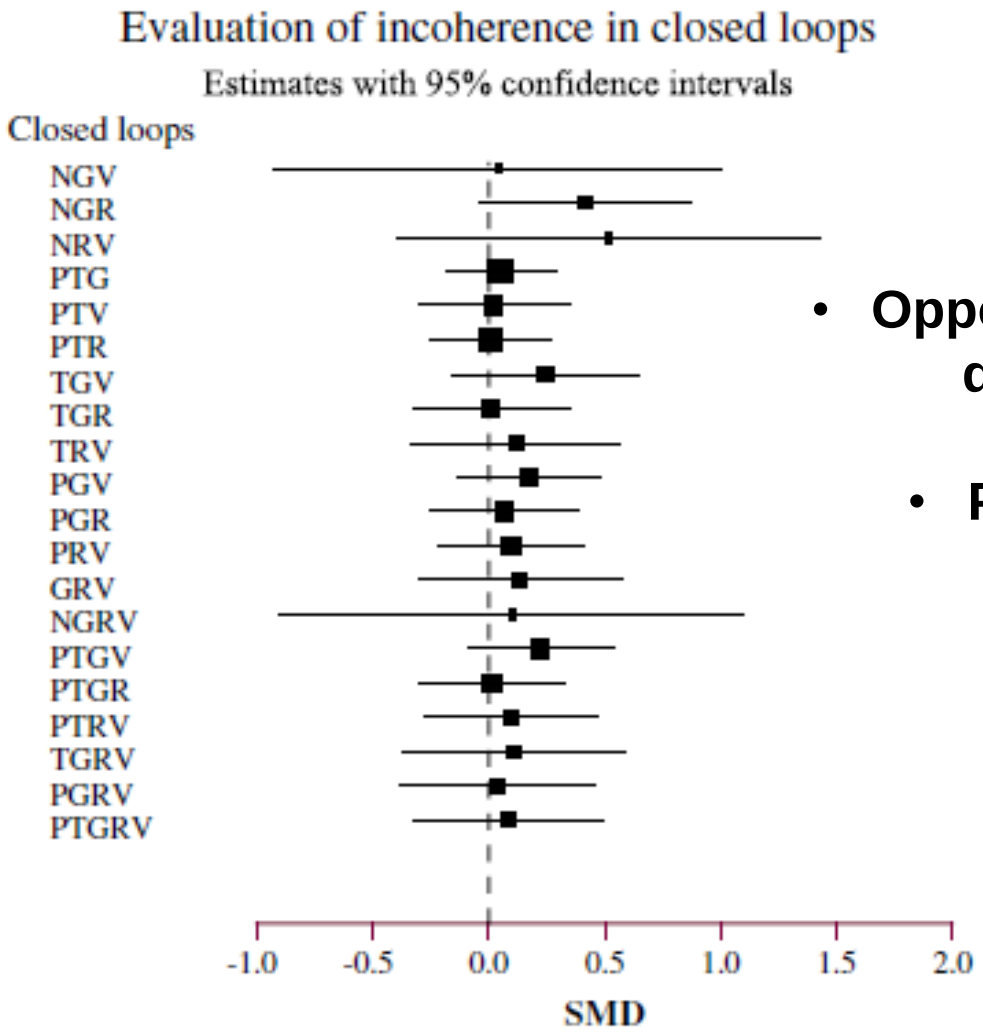
Fig. 1. Network of comparisons of fluoride therapies (T, G, V, and R), placebo (P), and no treatment (N) showing numbers of trials in which each pairwise comparison has been made. The size of the lines and nodes is proportional to the amount of available information.

- Effect measure: standardised change in DMFS (decayed, missing, filled tooth surfaces)

Incoherence

- Direct estimate of effect of toothpaste compared with rinse: δ_{TR}^{D}
- Indirect estimate via common placebo comparator: $\delta_{TR}^{I} = \delta_{TP}^{D}$-$\delta_{RP}^{D}$
- Incoherence: $\phi = \delta_{TR}^{D}$-δ_{TR}^{I}
- In absence of multi-arm trials: $\widehat{\phi} = \widehat{\delta_{TR}^{D}}$- $\widehat{\delta_{TR}^{I}}$
- Variance: $\mathrm{var}(\widehat{\phi}) = var(\widehat{\delta_{TR}^{D}})$- $\mathrm{var}(\widehat{\delta_{TR}^{I}})$

Estimates of incoherence



- **Opportunity to observe incoherence depends on network structure**
- **Power to detect incoherence**

Potential Covariables

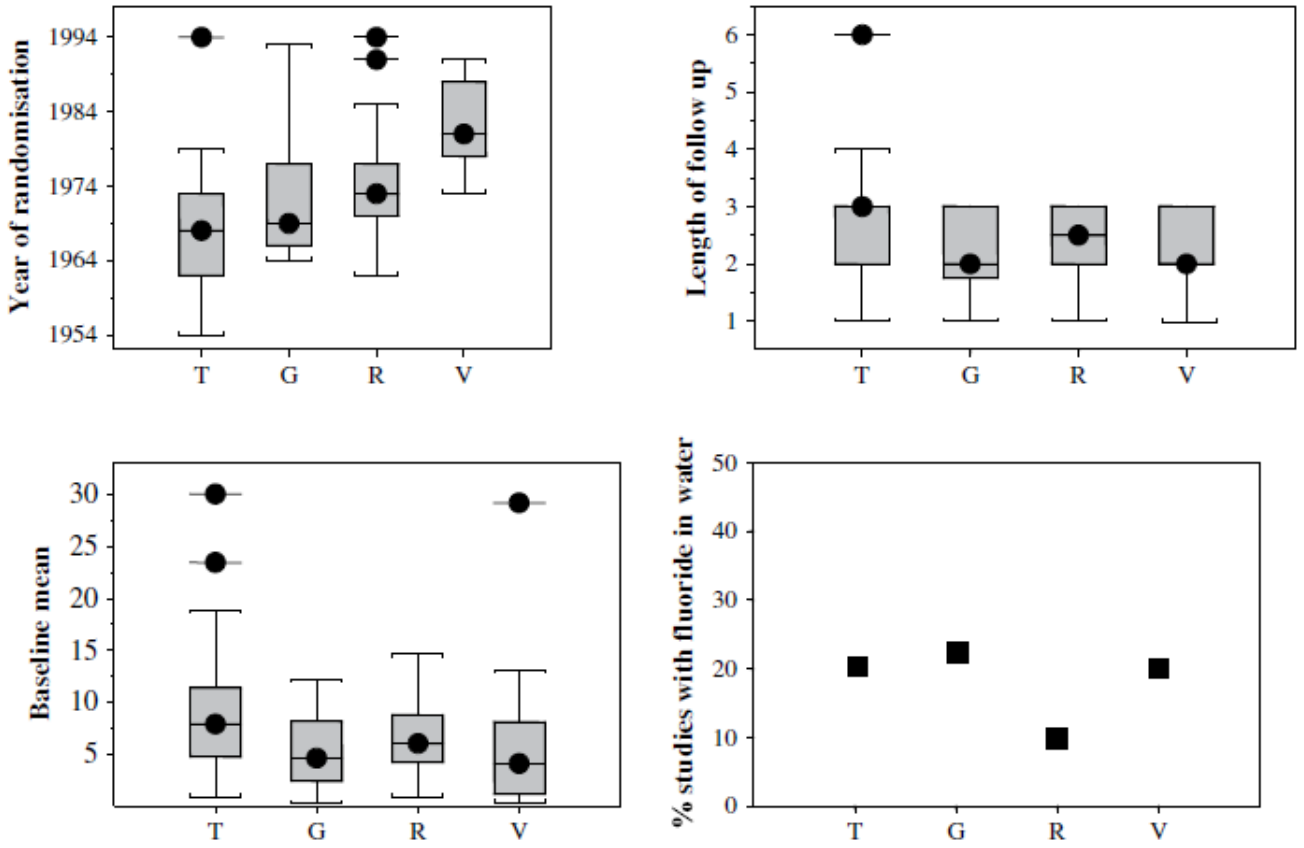


Fig. 4. Distribution of the possible confounding factors: box plots for year of randomization, length of follow-up, and baseline mean caries; percentage of the studies carried out in populations with fluoridation in the water.

Covariable Regression Model

$$\delta_{FC,i} = \delta'_{FC,i} + \beta(Year_i - Year_0)$$

Treatment effect for fluoride treatment vs control (placebo or not treatment) in year 0

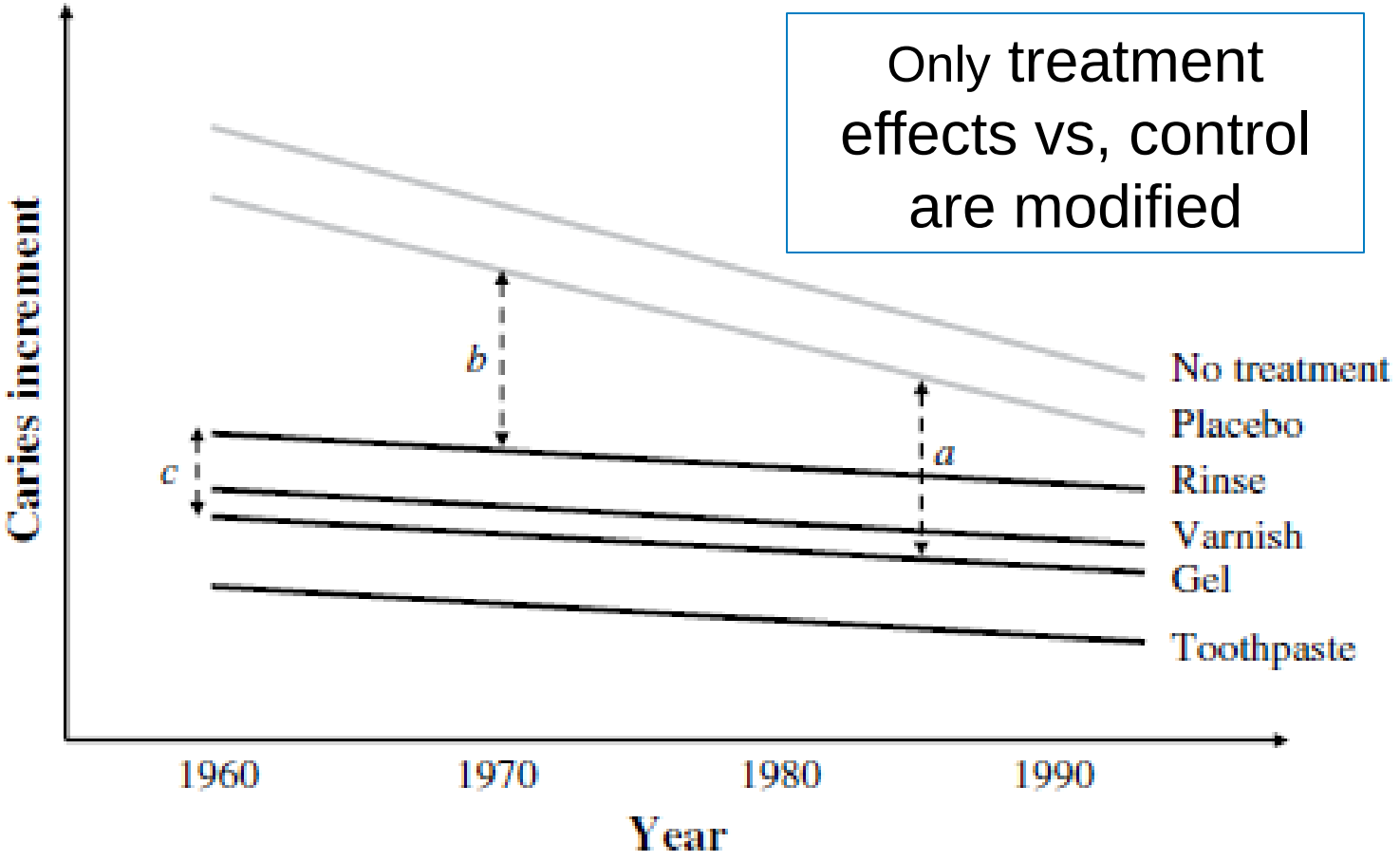
Covariable effects common across fluoride treatments

$$\delta'_{FC,i} = N(\delta'_{FC}, \tau^2_{FC})$$

Contrast dependent random treatment effects variance.

Effect of Adjustment

Not parameterisation invariant



Lumping and splitting: Different assumptions regarding the equivalence of Placebos

Table 4
Possible assumptions about placebo effects, and deviance information criteria (DIC) from network meta-analyses based on them

Model	Placebo assumption		Median τ	DIC
1	All placebos different	P_T, P_G, P_R, P_V	0.18	−80.7
2	Gel and varnish placebos equivalent	P_T, P_R, $P_G = P_V$	0.18	−81.0
3	Gel, varnish, and rinse placebos equivalent	P_T, $P_G = P_R = P_V$	0.18	−81.8
4	All placebos equivalent	$P_T = P_G = P_R = P_V$	0.18	−82.1
5	All placebos equivalent to no treatment	$N = P_T = P_G = P_R = P_V$	0.19	−80.5

Only possible as the network is well connected

STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:1861–1881
Published online 27 April 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3594

Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation

Nicola J. Cooper[1,*,†,‡], Alex J. Sutton[1,§], Danielle Morris[2], A. E. Ades[3,¶] and Nicky J. Welton[3,‖]

[1]*Department of Health Sciences, University of Leicester, Leicester LE1 7RH, U.K.*
[2]*Section of Epidemiology, Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, U.K.*
[3]*Academic Unit of Primary Health Care, Department of Community Based Medicine, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, U.K.*

SUMMARY

Mixed treatment comparison models extend meta-analysis methods to enable comparisons to be made between all relevant comparators in the clinical area of interest. In such modelling it is imperative that potential sources of variability are explored to explain both heterogeneity (variation in treatment effects between trials within pairwise contrasts) and inconsistency (variation in treatment effects between pairwise contrasts) to ensure the validity of the analysis.

The objective of this paper is to extend the mixed treatment comparison framework to allow for the incorporation of study-level covariates in an attempt to explain between-study heterogeneity and reduce inconsistency. Three possible model specifications assuming different assumptions are described and applied to a 17-treatment network for stroke prevention treatments in individuals with non-rheumatic atrial fibrillation.

The paper demonstrates the feasibility of incorporating covariates within a mixed treatment comparison framework and using model fit statistics to choose between alternative model specifications. Although such an approach may adjust for inconsistencies in networks, as for standard meta-regression, the analysis will suffer from low power if the number of trials is small compared with the number of treatment comparators. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: mixed treatment comparison; heterogeneity; atrial fibrillation

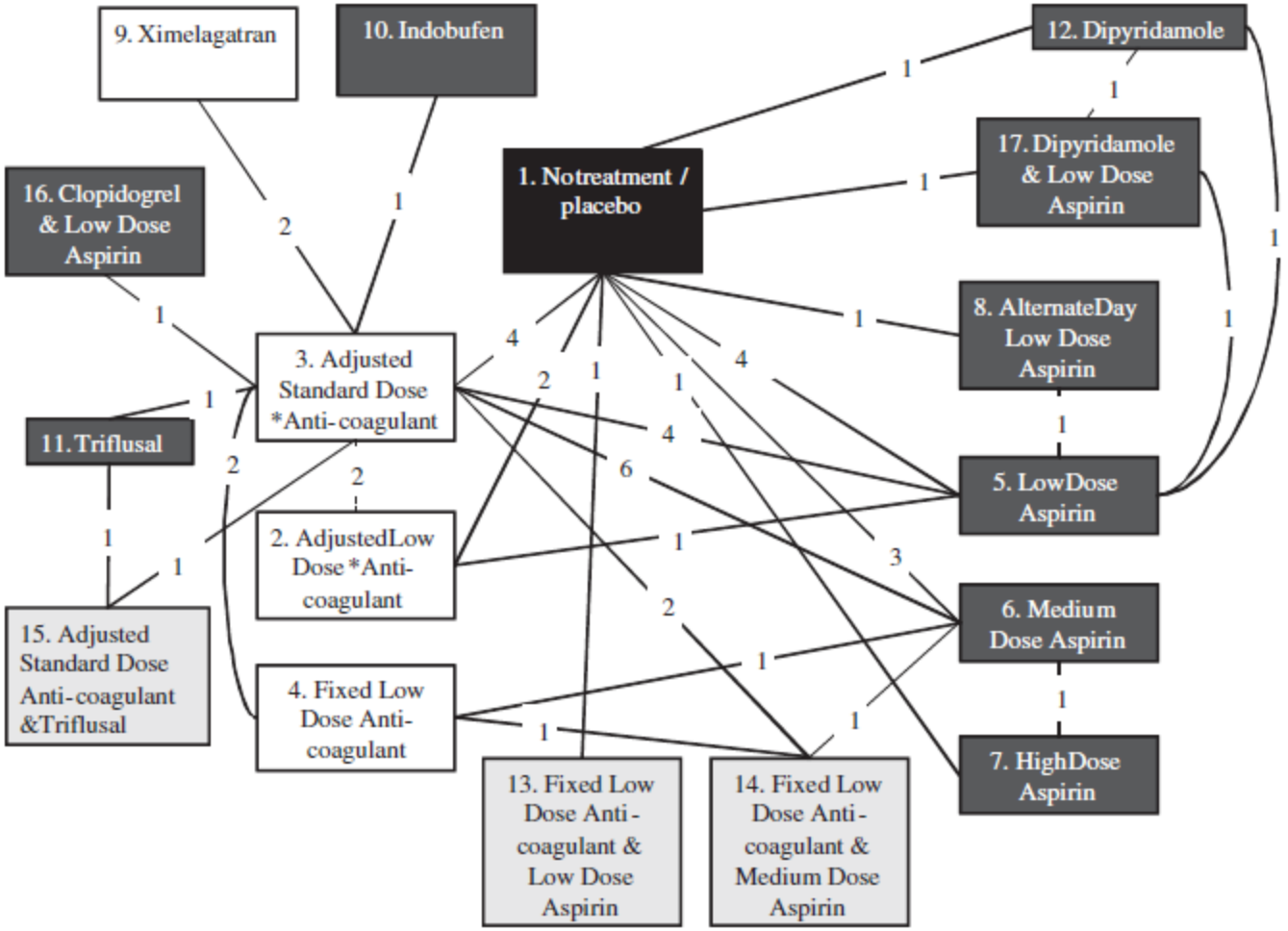


Figure 2. Full treatment network diagram for atrial fibrillation example (box shading indicates treatment class: white = Anti-coagulant, dark grey = Anti-platelet, light grey = Mixed (Anti-coagulant + Anti-platelet), black = placebo/no treatment). Each treatment strategy is a node in the network. The links between the nodes are trials or pairs of trial arms. The numbers along the link lines indicate the number of trials or pairs of trial arms for that link in the network. *In a number of trials clinicians were free in their choice of oral anti-coagulants; thus adjusted standard and low dose anti-coagulant treatment includes warfarin, phenprocoumon and acenocoumarol among others.

Covariable Regression Models

- Different regression coefficient for each comparison

$$\delta_{bk} = d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab}).X_j$$

- Exchangeable regression co-efficients

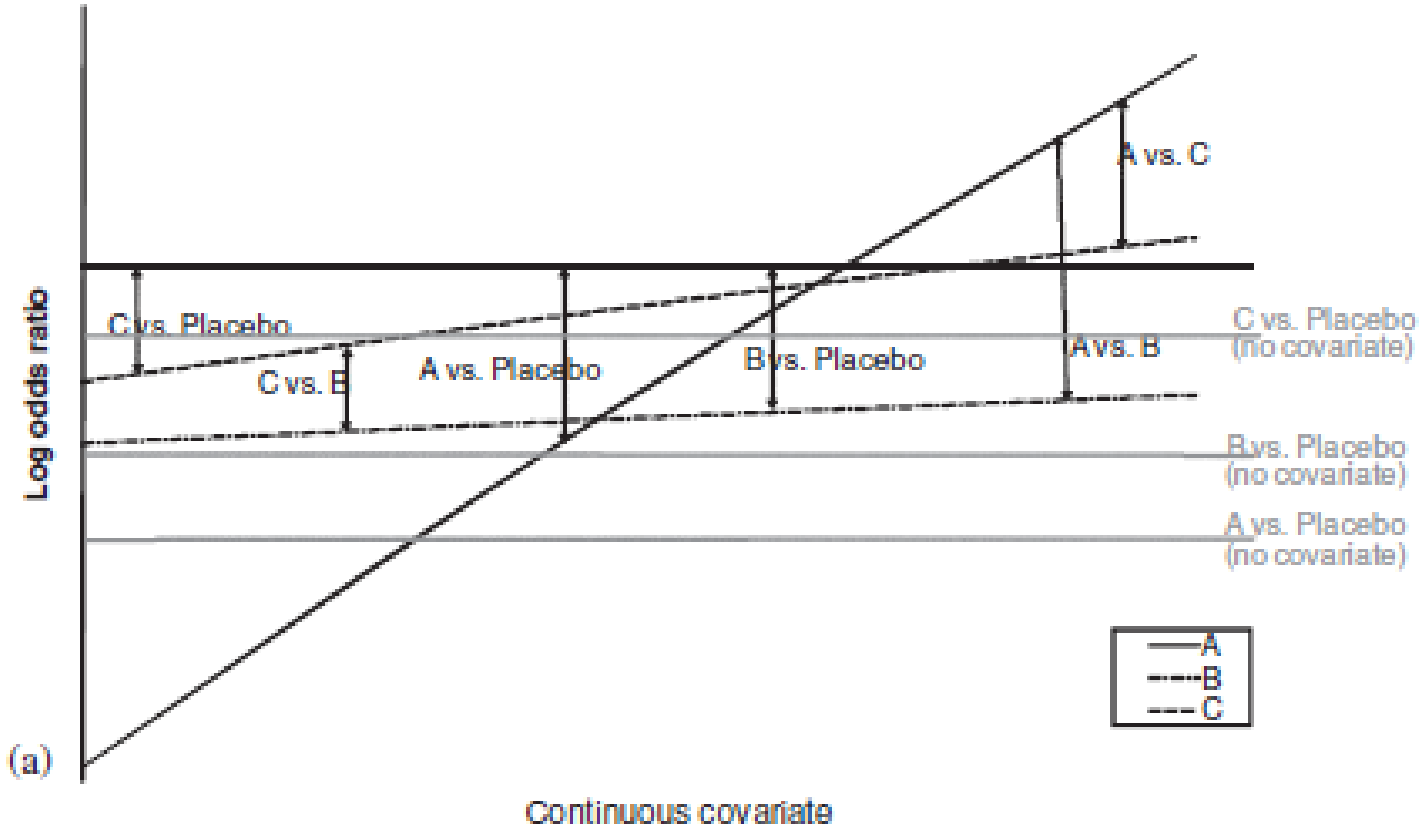
$$\delta_{bk} = d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab}).X_j$$

$$\beta_{Ak} \sim N\left(\beta, \sigma^2{}_{\beta}\right)$$

- Common regression co-efficient

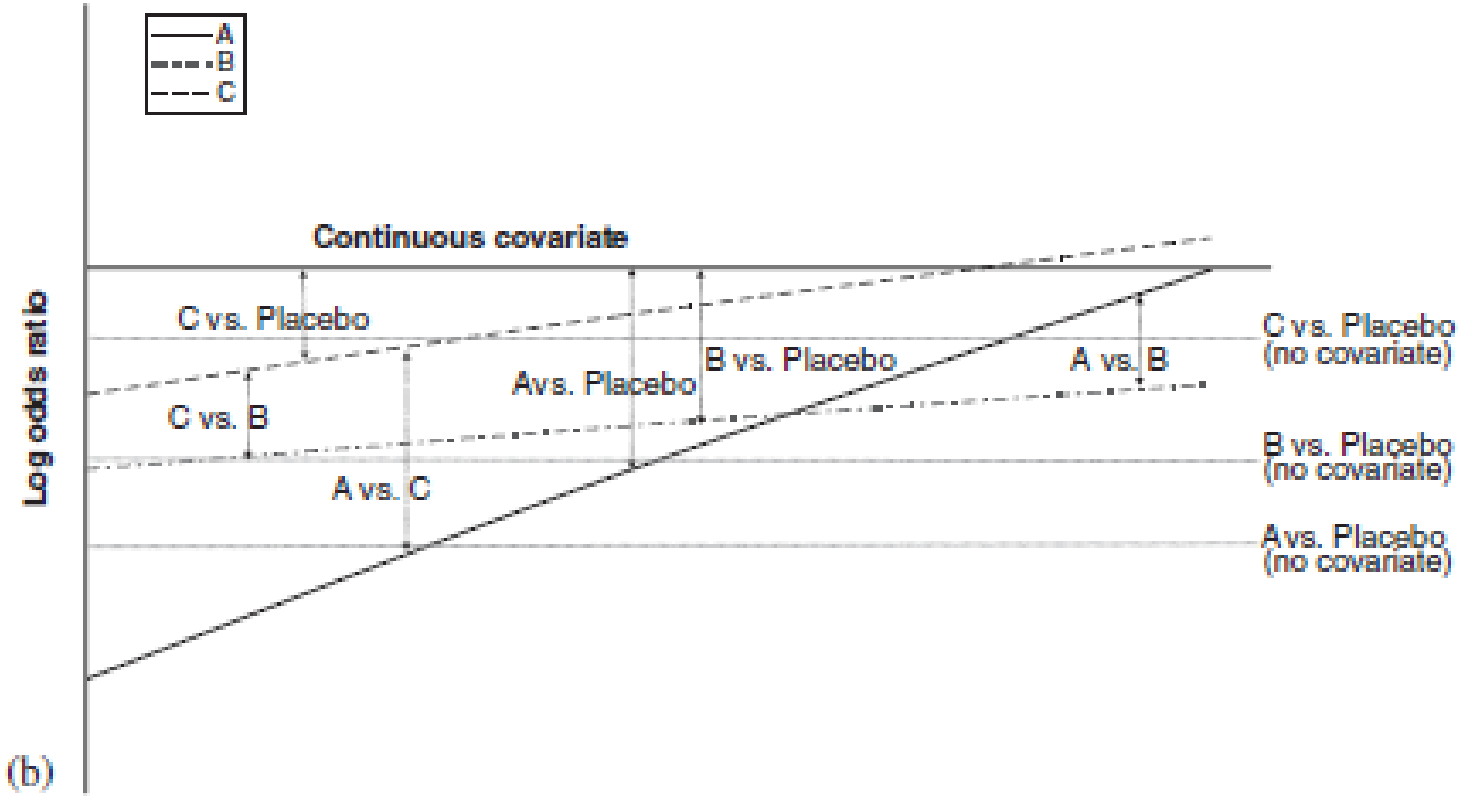
$$\delta_{bk} = d_{Ak} - d_{Ab} + \beta.X_j$$

Different Co-efficients

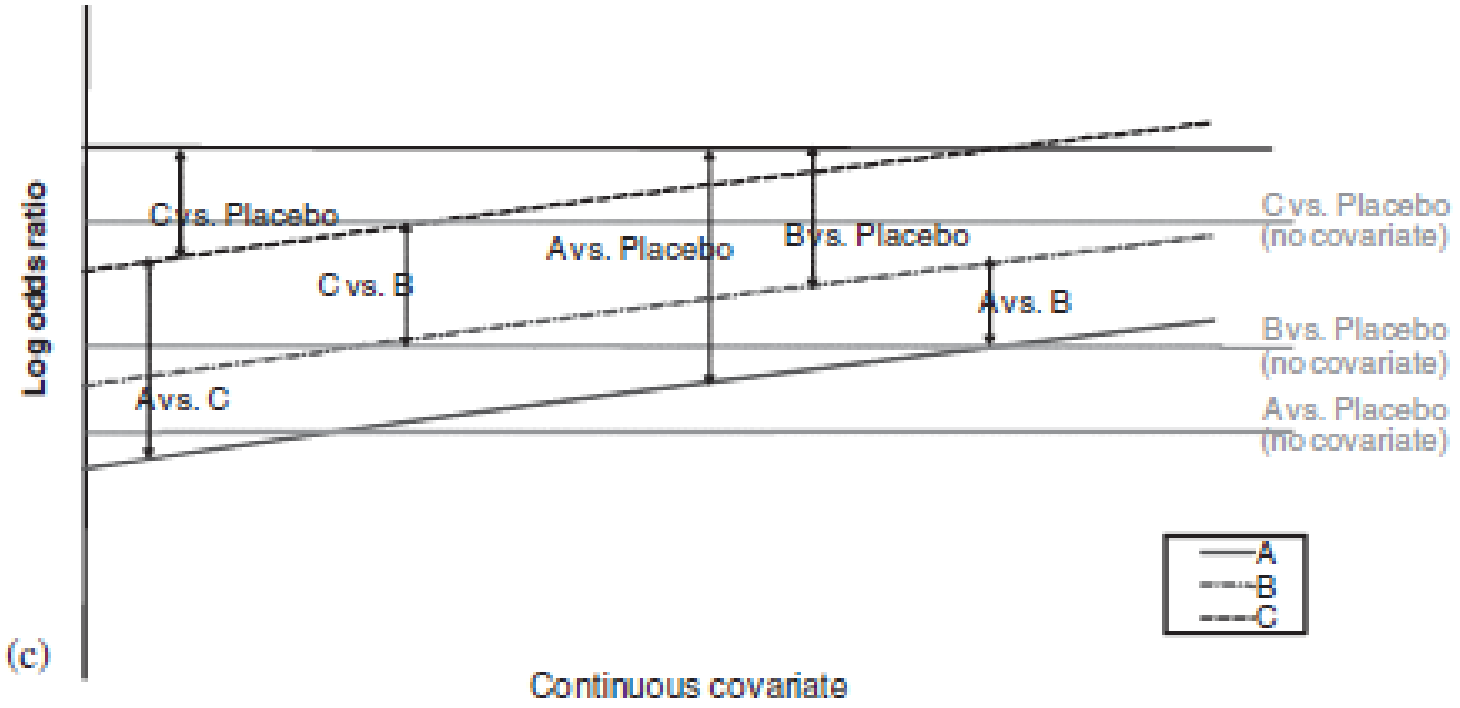


Exchangeable Co-efficients

Shrinkage



Common Co-efficient



Even more models

Model 1: Mixed treatment comparison with no covariates.

Model 3(a): Exchangeable *treatment x covariate* effects across all treatments.

Model 3(b): Exchangeable *treatment x covariate* effects within treatment class (i.e. Anti-coagulant, Anti-platelet, Mixed treatment groups having different interaction mean effects) with a common variance across class random effects.

Model 3(c): Exchangeable *treatment x covariate* effects within treatment classes and with the class interaction distributions having different variances.

Model 4(a): The *treatment x covariate* effects are identical across all treatment regimes.

Model 4(b): The *treatment x covariate* effects are identical within treatment classes (i.e. Anti-coagulant, Anti-platelet, Mixed (Anti-coagulant + Antiplatelet)).

Model Comparison

Table II. Model fit statistics.

Model fit		Model (1) No Covariates	Model 3(a) Exchangeable *treatment x covariate* effects	Model 3(b) Exchangeable *treatment x covariate* effects by class	Model 3(c) Exchangeable *treatment x covariate* effects by class—diff vars.	Model 4(a) *Same treatment x covariate* effects	Model 4(b) *Same treatment x covariate* effects by class
Residual deviance*	$\bar{D}$	60.22	58.74	57.71	57.72	59.97	58.74
Effective number of parameters	pD	48.35	48.26	48.88	49.20	49.81	48.25
Deviance Information Criteria	DIC	108.57	106.99	106.60	106.92	109.78	106.99

*Compared with 60 unconstrained data points.

Results

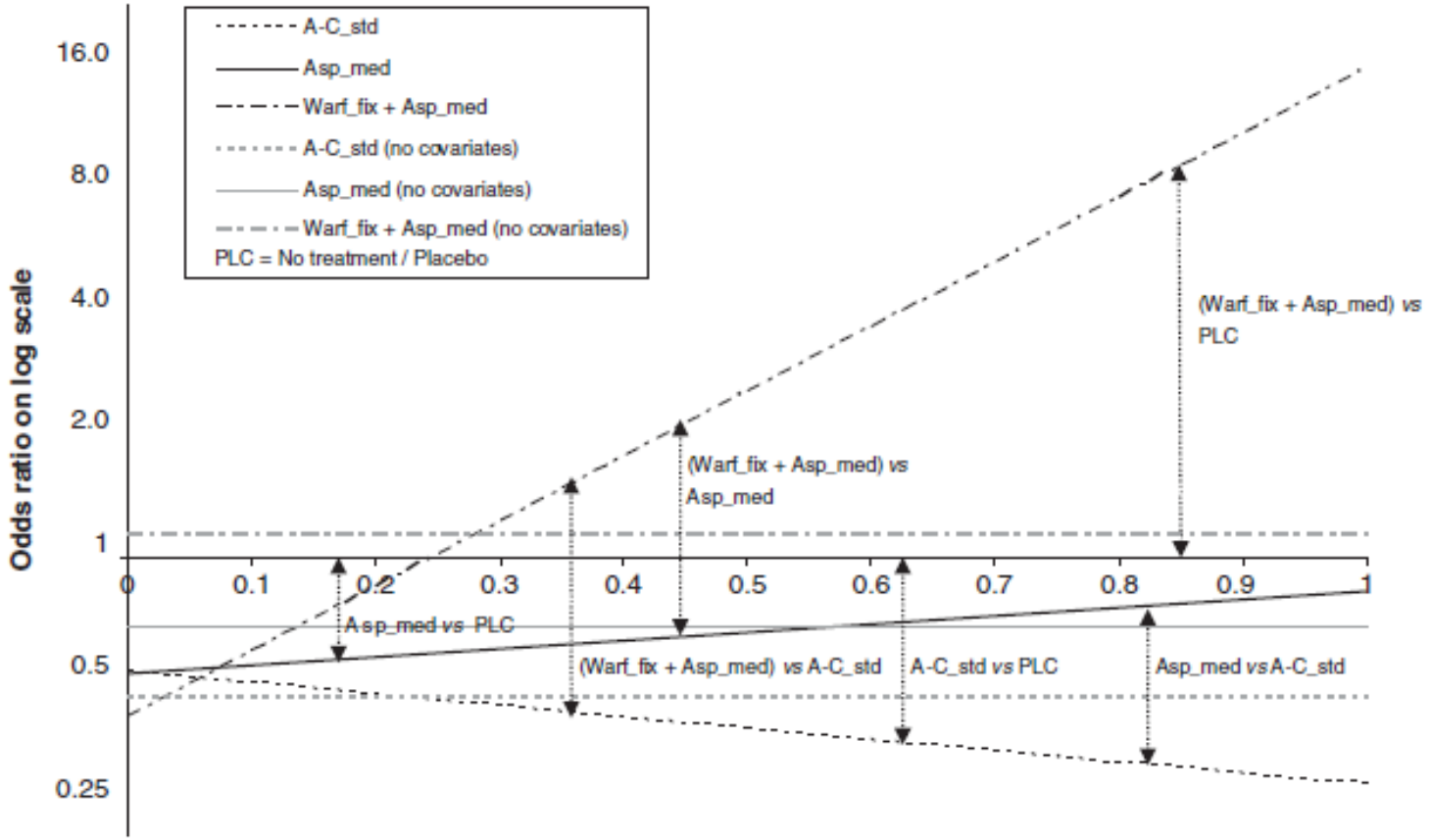


Figure 3. Mixed treatment comparison results adjusting for the proportion of patients with previous stroke or transient ischemic attack—exchangeable treatment x covariate effect within treatment class with same standard deviation (Key: A–C_std = standard dose anti-coagulant, Asp_med = medium dose aspirin, Warf_fix + Asp_med = fixed dose warfarin + medium dose aspirin).

Special Issue Paper

Research Synthesis Methods

Received 22 June 2011, Revised 11 May 2012, Accepted 15 May 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1044

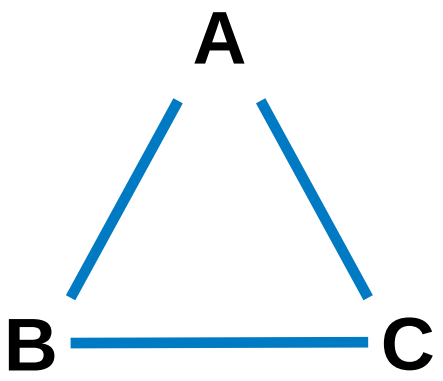
Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies‡

J. P. T. Higgins,[a,b*†] D Jackson,[a] J. K. Barrett,[a] G Lu,[c] A. E. Ades[c] and I. R. White[a]

Meta-analyses that simultaneously compare multiple treatments (usually referred to as network meta-analyses or mixed treatment comparisons) are becoming increasingly common. An important component of a network meta-analysis is an assessment of the extent to which different sources of evidence are compatible, both substantively and statistically. A simple indirect comparison may be confounded if the studies involving one of the treatments of interest are fundamentally different from the studies involving the other treatment of interest. Here, we discuss methods for addressing inconsistency of evidence from comparative studies of different treatments. We define and review basic concepts of heterogeneity and inconsistency, and attempt

Definitions of inconsistency

- Consistency ($\delta_{AC} = \delta_{AB} + \delta_{BC}$)
- Heterogeneity
  - Variation within a comparison ($\delta_{AB,i} \neq \delta_{AB,j}$)
- Loop inconsistency
  - Differences in treatment
    effect modifiers between comparisons
  - Average treatment effects
    not consistent around loop

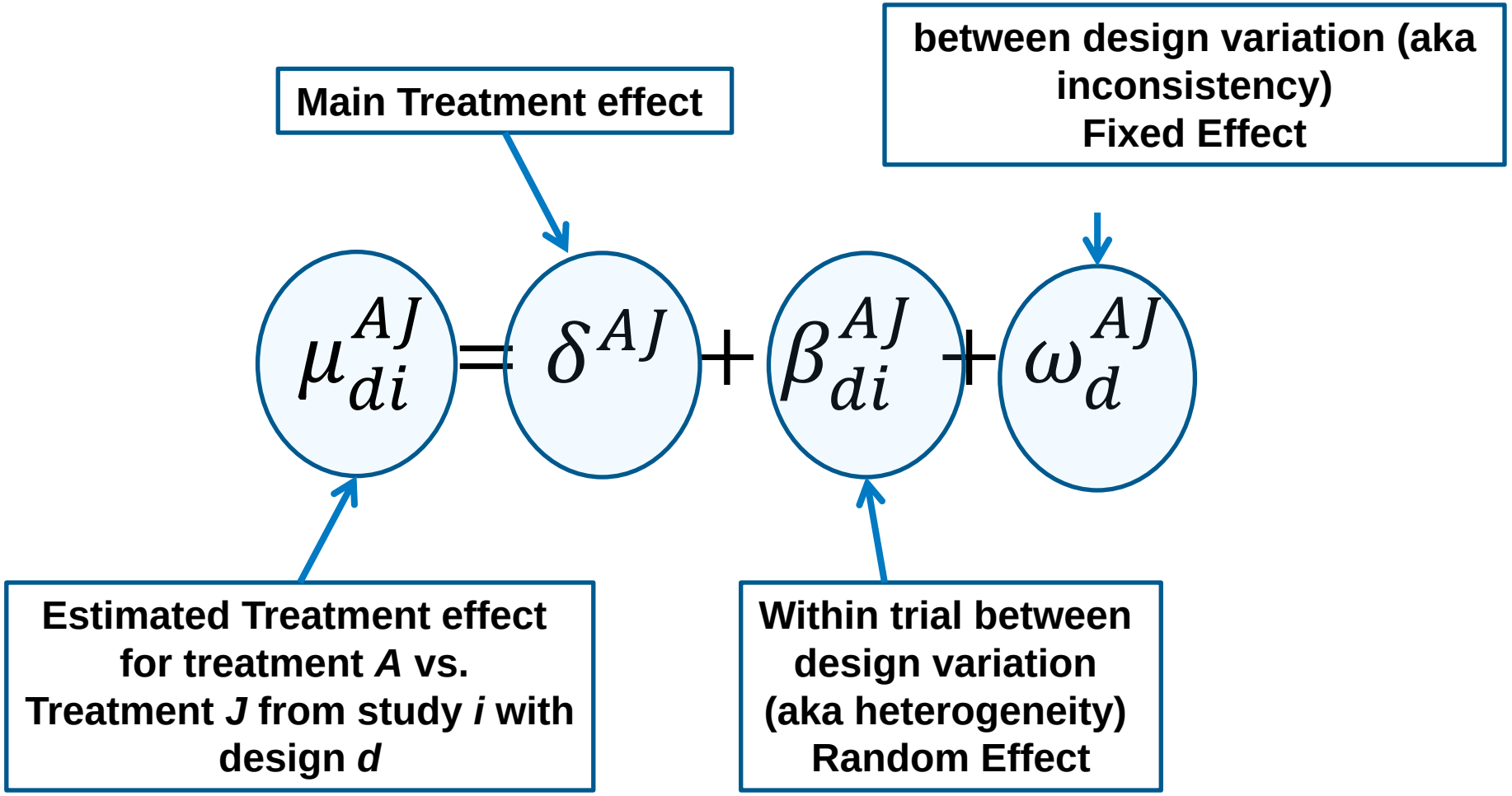


- Design inconsistency
  - Treatment effects vary by study design (design=comparator set)

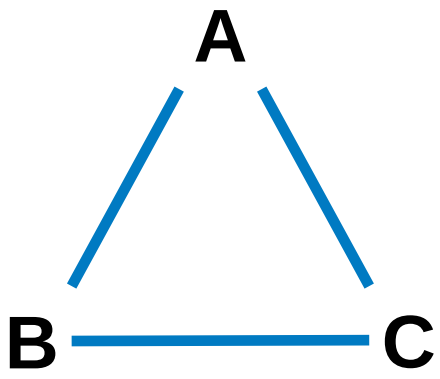
Observations

- Heterogeneity
  - Reflects presence of treatment effect modifiers
  - Only type of inconsistency if all trials include all comparators
- Loop inconsistency
  - Arises because of missing comparators
- Design inconsistency
  - Design = study level covariate that modifies treatment effects.
  - Special case of heterogeneity
  - Not distinguishable from loop inconsistency if all studies two armed
  - 25% of comparative studies have >2 arms

Consistency Model



An Example



	Treatment		
Trial Design	A	B	C
ABC	Ref	δ^{AB}	δ^{AC}
AB	Ref	$\delta^{AB} + \omega_2^{AB}$	-
AC	Ref	-	$\delta^{AC} + \omega_3^{AC}$
BC	Ref	δ^{AB}	$\delta^{AC} + \omega_4^{AC}$

American Journal of Epidemiology
© The Author 2009. Published by the Johns Hopkins Bloomberg School of Public Health.
All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.
Vol. 169, No. 9
DOI: 10.1093/aje/kwp014
Advance Access publication March 3, 2009

Practice of Epidemiology

Mixed Treatment Comparison Meta-Analysis of Complex Interventions: Psychological Interventions in Coronary Heart Disease

Nicky J. Welton, D. M. Caldwell, E. Adamopoulos, and K. Vedhara

Initially submitted March 14, 2008; accepted for publication January 12, 2009.

Meta-analyses of psychological interventions typically find a pooled effect of "psychological intervention" compared with usual care. This answers the research question, "Are psychological interventions in general effective?" In fact, psychological interventions are usually complex with several different components. The authors propose that mixed treatment comparison meta-analysis methods may be a valuable tool when exploring the efficacy of interventions with different components and combinations of components, as this allows one to answer the research question, "Are interventions with a particular component (or combination of components) effective?" The authors illustrate the methods using a meta-analysis of psychological interventions for patients with coronary heart disease for a variety of outcomes. The authors carried out systematic literature searches to update an earlier Cochrane review and classified components of interventions into 6 types: usual care, educational, behavioral, cognitive, relaxation, and support. Most interventions were a combination of these components. There was some evidence that psychological interventions were effective in reducing total cholesterol and standardized mean anxiety scores, that interventions with behavioral components were effective in reducing the odds of all-cause mortality and nonfatal myocardial infarction, and that interventions with behavioral and/or cognitive components were associated with reduced standardized mean depression scores.

Bayesian inference; coronary disease; Markov chain Monte Carlo; meta-analysis

Abbreviations: DIC, deviance information criterion; SMD, standardized mean difference.

Available Trials

Table 2. Intervention Components by Study Arm[a]

Intervention	Total No. of Arms	No. of Trial Arms by Outcome With Intervention							
		All-Cause Mortality	Cardiac Mortality	Nonfatal Myocardial Infarction	Total Cholesterol	Systolic Blood Pressure	Diastolic Blood Pressure	Depression	Anxiety
Usual care only	51 (7)	36 (5)	15 (2)	22 (4)	14	9	9	19 (3)	14
Educational	3 (1)	3 (1)	1	1	1	1	1	1 (1)	0
Behavioral	6 (2)	6 (1)	4 (1)	5 (2)	2	0	0	1	0
Cognitive	9 (5)	7 (2)	5 (3)	6 (4)	2	2	2	5 (1)	3
Support	1	1	0	0	1	1	1	0	1
Educational + behavioral	3	2	1	2	0	1	1	2	1
Educational + cognitive	5 (4)	5 (4)	1	2 (1)	1	1	1	4	2
Educational + relaxation	2	2	0	0	0	0	0	1	1
Behavioral + cognitive	4	2	1	1	2	0	0	0	0
Behavioral + relaxation	1	1	1	1	0	0	0	0	0
Cognitive + relaxation	2	1	0	1	1	1	1	1	1
Cognitive + support	1	1	0	1	0	0	0	0	0
Educational + behavioral + cognitive	2	1	1	1	0	0	0	1	0
Educational + behavioral + relaxation	3 (1)	3 (1)	0	3 (1)	1	0	0	2	2
Educational + behavioral + support	1	2	1	0	1	1	1	0	0
Educational + cognitive + relaxation	2 (1)	2 (1)	1	1	0	0	0	0	0
Behavioral + cognitive + relaxation	1	0	0	0	0	0	0	1	1
Behavioral + cognitive + support	1	2	0	1	1	1	1	1	1
Educational + behavioral + cognitive + relaxation	2	0	0	0	1	0	0	0	1

[a] Numbers in parentheses indicate the number of arms from 3-arm trials.

Models for Intervention Effects

- Single effect model (*all interventions created equal*)

$$d_k = d_.$$

- Additive main effects (*whole = sum of individual parts*)

$$d_k = d_{EDU} \times I_{k \supset EDU} + d_{BEH} \times I_{k \supset BEH} + \cdots$$

- 2-way interaction model (*whole = sum of pairs*)

$$d_k = d_{EDU} \times I_{k \supset EDU} + d_{BEH} \times I_{k \supset BEH} + d_{EDU*BEH} \times I_{\{k \supset EDU,BEH\}} + \cdots$$

- Full interaction model (*each intervention is unique*)

$$d_k = d_k$$

Model Comparison

Table 3. Deviance Information Criterion to Compare Models 1–4 for Each of the Outcomes Measures, Assuming a Correlation of 0.5 Between Pre- and Postmeasures for the Continuous Outcomes[a]

Outcome	Deviance Information Criterion[b]			
	Model 1 (Single Effect)	Model 2 (Additive Main Effects)	Model 3 (2-Way Interaction)	Model 4 (Full Interaction)
All-cause mortality	361.1	360.6	360.9	362.9
Cardiac mortality	160.8 (147.6)	161.2 (150.4)	157.2 (151.7)	157.4 (151.5)
Nonfatal myocardial infarction	243.7	241.0	247.2	244.4
Total cholesterol	−21.0	−20.0	−18.7	−18.5
Systolic blood pressure	86.3	87.0	87.7	87.6
Diastolic blood pressure	71.0	70.7	70.5	70.5
Depression	121.9	123.5	121.6	123.2
Anxiety	72.4	78.9	82.0	82.1

[a] The numbers in parentheses for cardiac mortality are obtained after omitting study 5 (Cowan et al. *Nurs Res*. 2001;50(2):68–76 (15)). Note that, in this example, there is little power to detect interaction effects (models 3 and 4).

[b] The sum of a measure of goodness of fit (posterior mean deviance) and a measure of model complexity (effective number of parameters).

Results

Table 4. Posterior Mean and 95% Credible Intervals for the Estimated Intervention Effect for Model 1 and Estimated Component Effects for Model 2[a]

Outcome	Summary	Model 1 (Single Effect)		Model 2 (Additive Main Effects)									
		d		d_{EDU}		d_{BEH}		d_{COG}		d_{REL}		d_{SUP}	
		Posterior Mean	95% Credible Interval	Posterior Mean	95% Credible Interval	Posterior Mean	95% Credible Interval	Posterior Mean	95% Credible Interval	Posterior Mean	95% Credible Interval	Posterior Mean	95% Credible Interval
All-cause mortality	Log-odds ratio	−0.14	−0.47, 0.15	0.29	−0.27, 0.85	−0.58	−1.13, −0.05	−0.01	−0.52, 0.45	−0.38	−1.16, 0.37	0.21	−0.66, 1.06
Cardiac mortality[b]	Log-odds ratio	−0.16	−0.44, 0.07	0.27	−0.46, 0.98	−0.34	−1.00, 0.30	−0.33	−0.83, 0.03	0.03	−1.49, 1.53	0.10	−1.12, 1.31
Nonfatal myocardial infarction	Log-odds ratio	−0.35	−0.65, −0.10	−0.16	−0.71, 0.34	−0.64	−1.13, −0.16	−0.09	−0.41, 0.28	−0.005	−0.61, 0.57	−1.49	−3.42, 0.19
Total cholesterol, mmol/L	Mean difference	−0.32	−0.50, −0.13	−0.13	−0.71, 0.42	−0.14	−0.60, 0.33	−0.29	−0.71, 0.13	0.49	−0.23, 1.24	−0.05	−0.70, 0.61
Systolic blood pressure, mm Hg	Mean difference	−1.21	−4.24, 2.33	−2.81	−12.84, 7.18	5.53	−8.61, 19.78	−0.95	−9.13, 7.80	−0.07	−17.54, 16.50	−0.74	12.38, 11.63
Diastolic blood pressure, mm Hg	Mean difference	−1.37	−3.31, 0.62	−3.77	−10.42, 3.00	3.18	−6.61, 12.48	0.89	−4.87, 6.44	−2.39	−14.00, 9.43	−0.85	−8.74, 7.40
Depression	SMD	−0.23	−0.35, −0.11	−0.01	−0.24, 0.22	−0.26	−0.55, 0.02	−0.24	−0.42, −0.06	0.08	−0.20, 0.34	0.57	−0.07, 1.21
Anxiety	SMD	−0.15	−0.29, −0.04	−0.19	−0.49, 0.14	−0.02	−0.37, 0.34	−0.12	−0.37, 0.10	0.02	−0.31, 0.34	−0.04	−0.39, 0.38

Abbreviations: SMD, standardized mean difference; subscript abbreviations: BEH, behavioral intervention; COG, cognitive intervention; EDU, educational intervention; REL, relaxation intervention; SUP, psychosocial support intervention.

[a] Results are shown for the relevant summary measure for each of the outcome measures, assuming a correlation of 0.5 between pre- and postmeasures for the continuous outcomes.

[b] Results presented for the cardiac mortality outcome omit study 5 (Cowan et al. *Nurs Res.* 2001;50(2):68–76 (15)).

Probability of being most effective

Table 5. Proportion of Simulations From Model 2 (Additive Main Effects) in Which Each Component Was the Most Effective (Had the Lowest Log-Odds Ratio, Mean Difference, or Standardized Mean Difference as Appropriate) for Each of the Outcome Measures, Assuming a Correlation of 0.5 Between Pre- and Postmeasures for the Continuous Outcomes

Outcome	Probability of Being Most Effective					
	Usual Care	Educational	Behavioral	Cognitive	Relaxation	Psychosocial Support
All-cause mortality	0.000	0.009	0.614	0.024	0.316	0.038
Cardiac mortality[a]	0.000	0.022	0.398	0.207	0.239	0.135
Nonfatal myocardial infarction	0.000	0.010	0.226	0.001	0.007	0.756
Total cholesterol	0.000	0.194	0.212	0.420	0.012	0.162
Systolic blood pressure	0.004	0.325	0.063	0.195	0.226	0.187
Diastolic blood pressure	0.001	0.465	0.056	0.043	0.317	0.117
Depression	0.000	0.039	0.515	0.423	0.015	0.008
Anxiety	0.000	0.483	0.142	0.176	0.075	0.125

[a] Results presented for the cardiac mortality outcome omit study 5 (Cowan et al. *Nurs Res.* 2001;50(2):68–76 (15)).

STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:5620–5639
Published online 4 August 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3377

Mixed treatment comparison with multiple outcomes reported inconsistently across trials: Evaluation of antivirals for treatment of influenza A and B

N. J. Welton[1,*,†], N. J. Cooper[2], A. E. Ades[1], G. Lu[1] and A. J. Sutton[2]

[1]*Academic Unit of Primary Health Care, Department of Community Based Medicine, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, U.K.*
[2]*Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, U.K.*

SUMMARY

We present a mixed treatment meta-analysis of antivirals for treatment of influenza, where some trials report summary measures on at least one of the two outcomes: time to alleviation of fever and time to alleviation of symptoms. The synthesis is further complicated by the variety of summary measures reported: mean time, median time and proportion symptom free at the end of follow-up. We compare several models using the deviance information criteria and the contribution of different evidence sources to the residual deviance to aid model selection. A Weibull model with exchangeable treatment effects that are independent for each outcome but have a common random effect mean for the two outcomes gives the best fit according to these criteria. This model allows us to summarize treatment effect on two outcomes in a single summary measure and draw conclusions as to the most effective treatment. Amantadine and Oseltamivir were the most effective treatments, with the probability of being most effective of 0.56 and 0.37, respectively. Amantadine reduces the duration of symptoms by an estimated 2.8 days, and Oseltamivir 2.6 days, compared with placebo. The models provide flexible methods for synthesis of evidence on multiple treatments in the absence of head-to-head trial data, when different summary measures are used and either different clinical outcomes are reported or where the same outcomes are reported at different or multiple time points. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian methods; decision models; evidence synthesis; Markov chain Monte Carlo simulation; model criticism

Trial Data

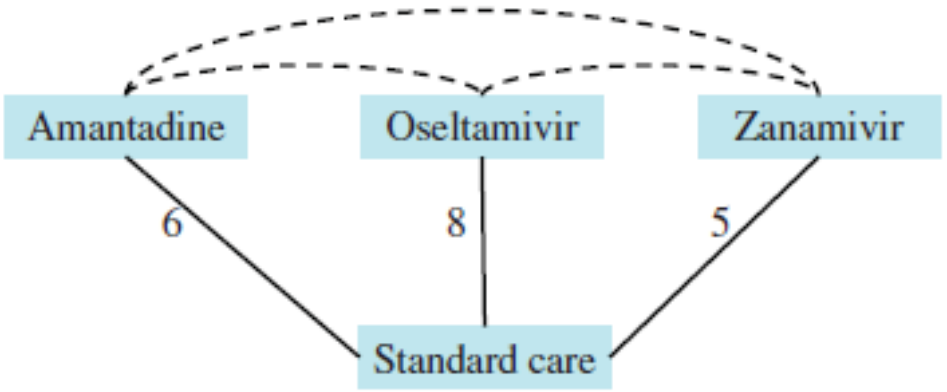


Figure 1. Treatment network diagram: each treatment strategy represents a node in the network. The numbers along the solid lines indicate the number of trials providing direct information for that link in the network. The dotted lines represent the indirect comparisons not evaluated in trials, but estimated via the model.

Disease Model

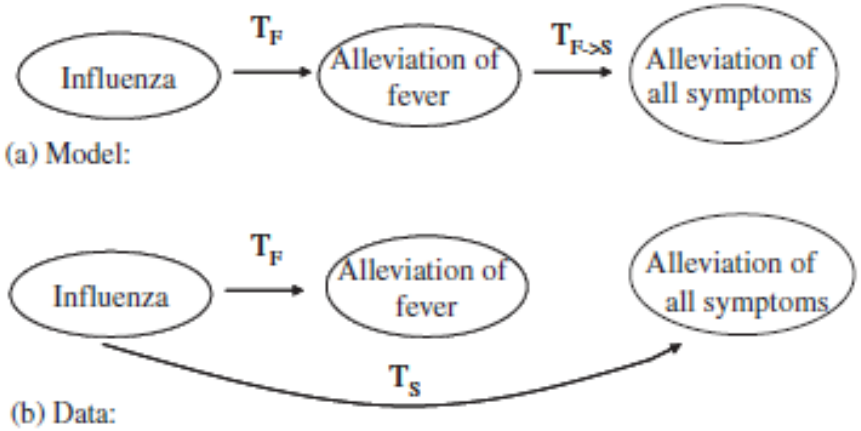


Figure 2. (a) The assumed disease progression model for influenza A and B, where T_F is the time to alleviation of fever, and $T_{F->S}$ is the time from alleviation of fever to alleviation of all symptoms. (b) The components of disease progression that we have data on. The data provide information directly on the transition between influenza and alleviation of fever, T_F, but there is only indirect information on the time from alleviation of fever to alleviation of all symptoms, $T_{F->S}$, i.e. the difference in time to alleviation of all symptoms, T_S, and time to alleviation of fever, T_F.

Time to event models (1)

Time to alleviation of fever:

$$T_{F,j,k} \sim Weibull(\alpha_{F,j,k}, \beta_{F,j,k})$$

Time to alleviation of symptoms:

$$T_{S,j,k} \sim Weibull(\alpha_{S,j,k}, \beta_{S,j,k})$$

Shape parameters constrained to be positive (decreasing hazard) and exchangeable between study, treatment and outcome

$$\alpha_{F,j,k}, \alpha_{S,j,k} \sim Beta(a, b)$$

Time to event models (2)

Symptoms alleviate after fever:

$$\beta_{S,j,k} \geq \beta_{F,j,k}$$

Placebo scale parameters unconstrained:

$$\log(\beta_{F,j,1}) = \mu_{F,j}$$

$$\log(\beta_{S,j,1}) = \mu_{F,j} + \gamma_j$$

$$\gamma_j + N(g, v)$$

Time to event models (2)

Symptoms alleviate after fever:

$$\beta_{S,j,k} \geq \beta_{F,j,k}$$

Placebo scale parameters unconstrained:

$$\log(\beta_{F,j,1}) = \mu_{F,j}$$

$$\log(\beta_{S,j,1}) = \mu_{F,j} + \gamma_j$$

$$\gamma_j + N(g, v)$$

Indirect comparison of treatment effects

log scale parameters transitive:

$$\log(\beta_{F,j,k}) = \mu_{F,j} + \delta_{F,j,k}$$

$$\log(\beta_{S,j,k}) = \mu_{F,j} + \gamma_j + \max(\delta_{S,j,k}, \delta_{F,j,k} - \gamma_j)$$

Random effects correlated

log scale parameters transitive:

$$\begin{pmatrix} \delta_{S,j,k} \\ \delta_{F,j,k} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} d_{S,k} \\ d_{F,k} \end{pmatrix}, \begin{pmatrix} \sigma_S^2 & \sigma_S \sigma_F \rho \\ \sigma_S \sigma_F \rho & \sigma_F^2 \end{pmatrix} \right)$$

Likelihoods

- *Median times (e.g. median time until alleviation of symptoms)*

$$y_{S,j,k}^{median} \sim N(\beta_{S,j,k}(ln2)^{1/\alpha_{S,j,k}}, (se_{S,j,k}^{median})^2)$$

- Count data (e.g. number individual symptom-free by day 28)

$$n_{S,j,k}^{28} \sim Bin\left(p_{S,j,k}^{28}, N_{j,k}\right)$$

$$p_{S,j,k}^{28} = \mathrm{E}(\Pr(T_{s,j,k} < 28 | y_{S,j,k}^{median})$$

A range of possible models

Table II. Model descriptions.

Model	Description
M1: Correlated treatment effects for fever and all symptoms, equal random effects variances	As described in Section 3.2, setting $\sigma_S = \sigma_F = \sigma$: $\begin{pmatrix}\delta_{S,j,k}\\ \delta_{F,j,k}\end{pmatrix} \sim N_2\left(\begin{pmatrix}d_{S,k}\\ d_{F,k}\end{pmatrix}, \begin{pmatrix}\sigma^2 & \sigma^2\rho\\ \sigma^2\rho & \sigma^2\end{pmatrix}\right)$
M2: Correlated treatment effects for fever and all symptoms, equal random effects means and variances	As described in Section 3.2, but setting $\sigma_S = \sigma_F = \sigma$ and $d_{S,k} = d_{F,k} = d_k$: $\begin{pmatrix}\delta_{S,j,k}\\ \delta_{F,j,k}\end{pmatrix} \sim N_2\left(\begin{pmatrix}d_k\\ d_k\end{pmatrix}, \begin{pmatrix}\sigma^2 & \sigma^2\rho\\ \sigma^2\rho & \sigma^2\end{pmatrix}\right)$
M3: Independent random effects for fever and all symptoms, with common means but unequal variances	As described in Section 3.2, but $\sigma_S \neq \sigma_F$, and setting $d_{S,k} = d_{F,k} = d_k$: $\begin{pmatrix}\delta_{S,j,k}\\ \delta_{F,j,k}\end{pmatrix} \sim N_2\left(\begin{pmatrix}d_k\\ d_k\end{pmatrix}, \begin{pmatrix}\sigma_S^2 & 0\\ 0 & \sigma_F^2\end{pmatrix}\right)$
M4: Independent random effects for fever and all symptoms, with common means and variances	As described in Section 3.2, but setting $\sigma_S = \sigma_F = \sigma$, $\rho = 0$ and $d_{S,k} = d_{F,k} = d_k$: $\begin{pmatrix}\delta_{S,j,k}\\ \delta_{F,j,k}\end{pmatrix} \sim N_2\left(\begin{pmatrix}d_k\\ d_k\end{pmatrix}, \begin{pmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right)$
M5: Treatment effect equal for fever and all symptoms for each study.	As described in Section 3.2, but setting: $\delta_{F,j,k} = \delta_{S,j,k} = \delta_{j,k}$; $\delta_{j,k} \sim N(d_k, \sigma^2)$
M4A: Shape parameters independent of study and treatment	As M4, but with $\alpha_S \sim \text{Beta}(2,2)$; $\alpha_F \sim \text{Beta}(2,2)$ priors
M4B: Shape parameters equal for fever and all symptoms, and independent of study and treatment	As M4, but with $\alpha_S = \alpha_F \sim \text{Beta}(2,2)$ prior
M4C: Exponential models for both fever and all symptoms	As M4, but with $\alpha_S = \alpha_F = 1$
M4D: Outcome and study-specific shape parameters	As M4, but with $\alpha_{S,j,k} \sim \text{Beta}(1,1)$; $\alpha_{F,j,k} \sim \text{Beta}(1,1)$

STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:3681–3699
Published online 6 February 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2831

Meta-analysis of mixed treatment comparisons at multiple follow-up times

G. Lu[1,*,†], A. E. Ades[1], A. J. Sutton[2], N. J. Cooper[2], A. H. Briggs[3] and D. M. Caldwell[1]

[1]*MRC Health Services Research Collaboration, University of Bristol, Bristol, U.K.*
[2]*Department of Health Sciences, University of Leicester, Leicester, U.K.*
[3]*Section of Public Health & Health Policy, University of Glasgow, Glasgow, U.K.*

SUMMARY

Mixed treatment comparisons (MTC) meta-analysis is a methodology for making inferences on relative treatment effects based on a synthesis of both direct and indirect evidence on multiple treatment contrasts. This is particularly useful in the context of cost-effectiveness analysis and medical decision making. Here, we extend these methods to a more complex situation where trials report results at one or more, different yet fixed, follow-up times. These methods are applied to an illustrative data set combining evidence on healing rates under six different treatments for gastro-esophageal reflux disease (GERD). A series of Bayesian hierarchical models based on piece-wise exponential hazards is developed that borrow strength across the MTC networks and also across time points. These include models for absolute and relative treatment effects, models with fixed or random effects over time, random walk models, and models with homogeneous or heterogeneous between-trials variation. The deviance information criterion (DIC) is used to guide model development and selection. Models for absolute treatment effects generate materially different rankings of the treatments than models that separate the trial-specific baselines from the relative treatment effects. The extent of between-trials heterogeneity in treatment effects depends on treatment contrast. In discussion we note that models of this type have a very wide potential application. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian hierarchical model; MCMC; piece-wise exponential healing time; mixed treatment comparisons; multiple follow-up times; WinBUGS

Study Network

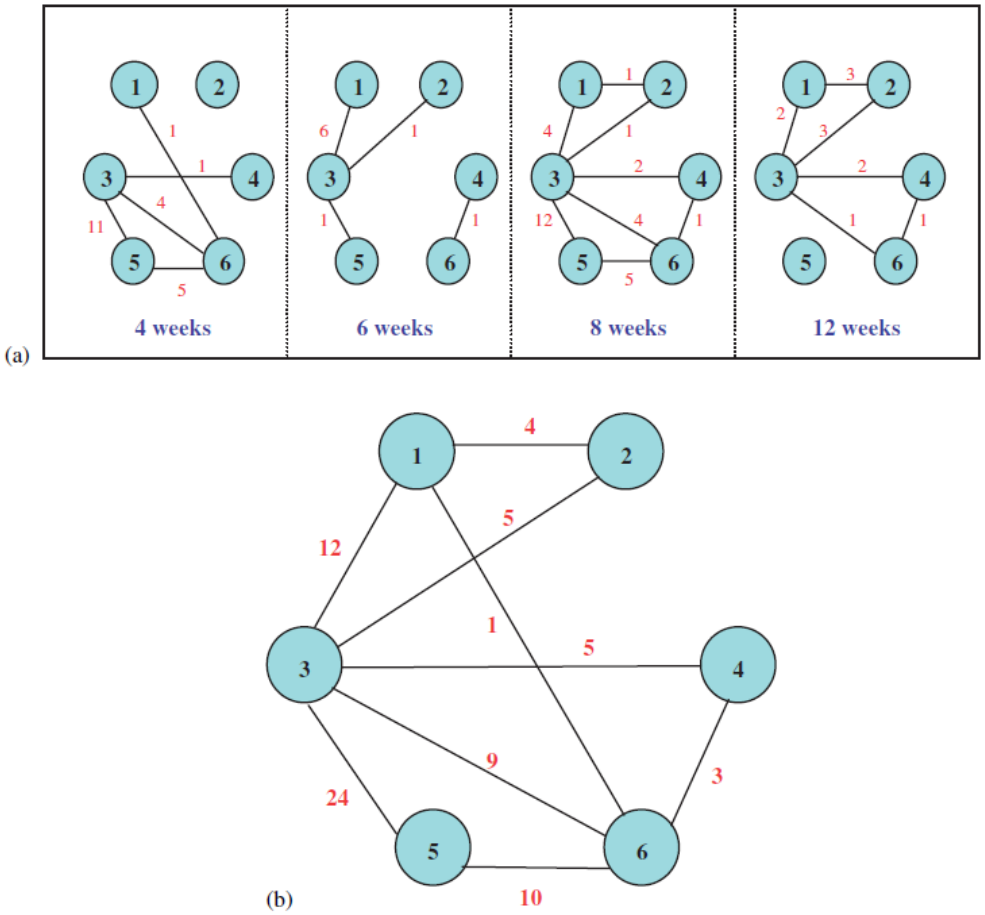
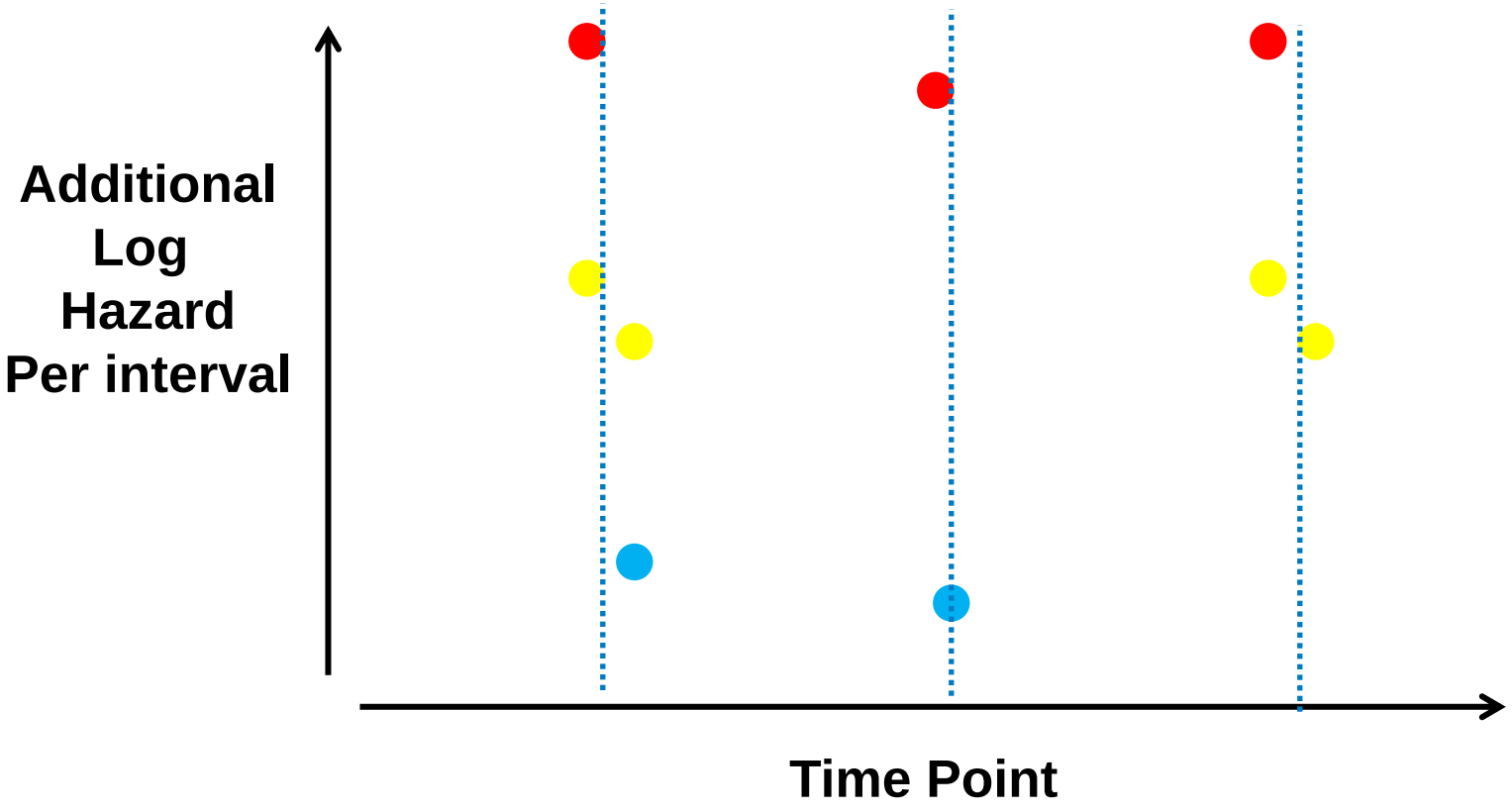


Figure 1. (a) Comparison networks at four follow-up times. The figure attached to each edge is the number of trials making corresponding comparison and (b) Comparison network aggregated over follow-up times. The figure attached to each edge is the number of trials making corresponding comparison.

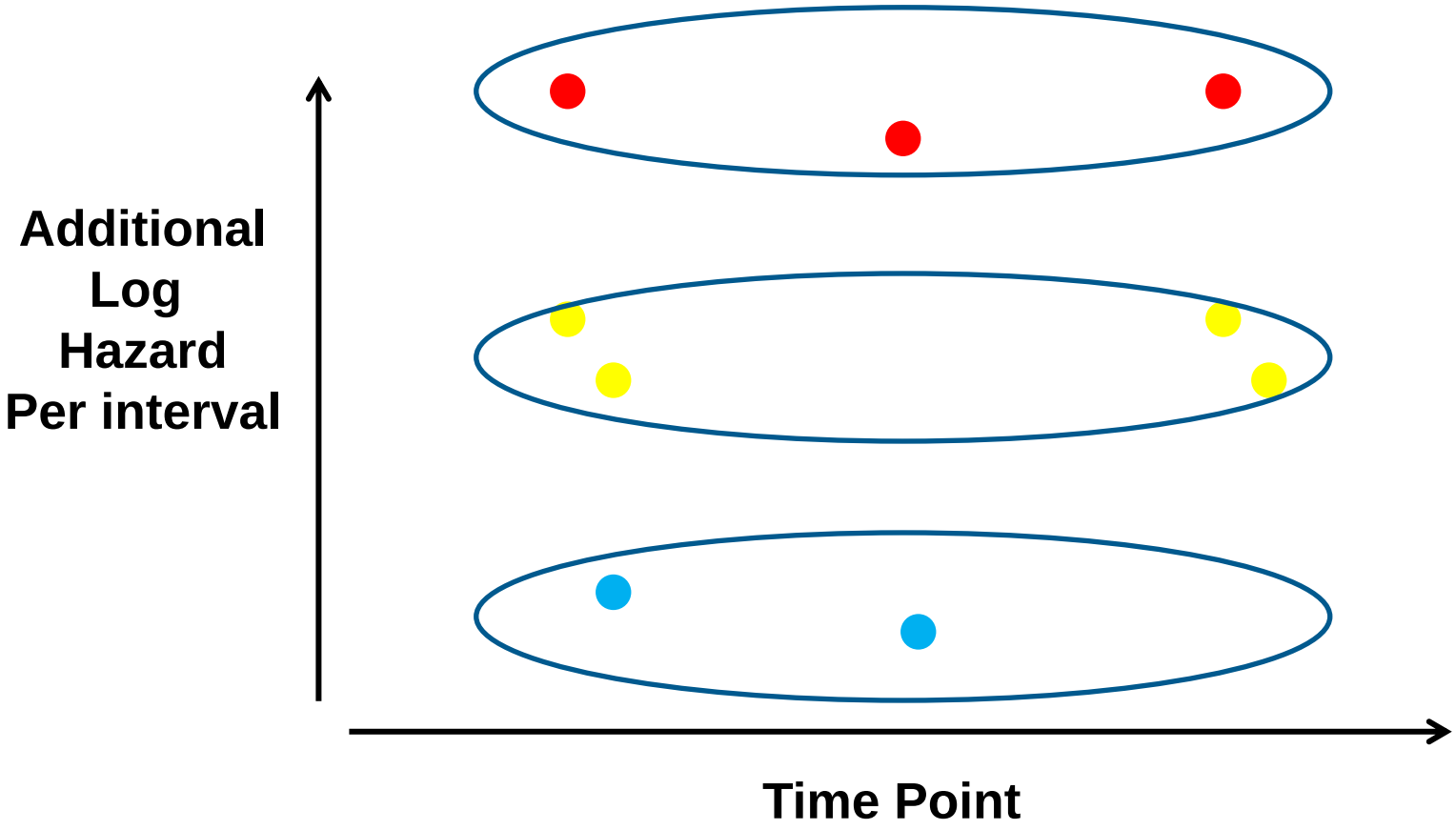
Model Options

- **Absolute log hazard (ALH) or relative log hazards (LHR); common across time periods**
- **Reference treatment response: fixed (F-B), mixed with random study by time period interaction (M-B), or random walk (RW-B); varies across time periods**
- **Treatment effects**
  - **fixed (F-LHR) or random (R-LHR); common across time periods**
  - **random (R-LHR(t)), or random walk (RW-LHR); varies across time periods**
- **Treatment effects variance: Homogeneous (Hom-V) or heterogeneous (Het-V) between treatments**

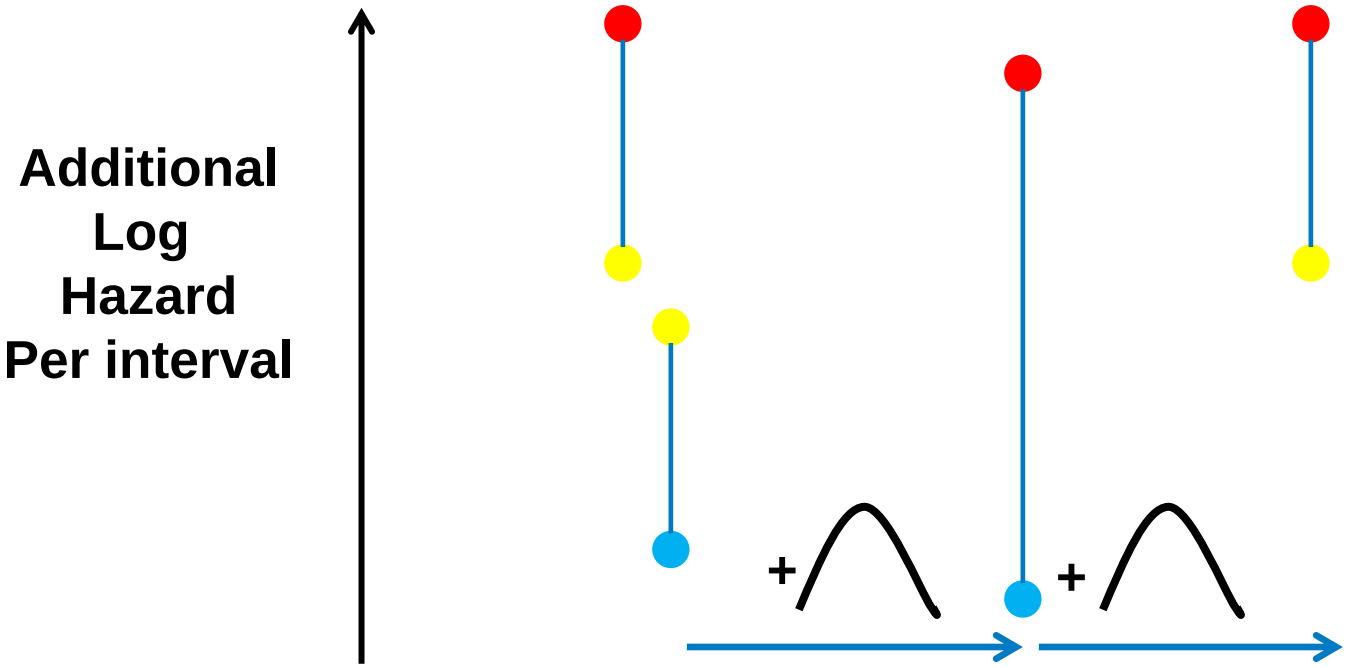
Illustrative Study Data



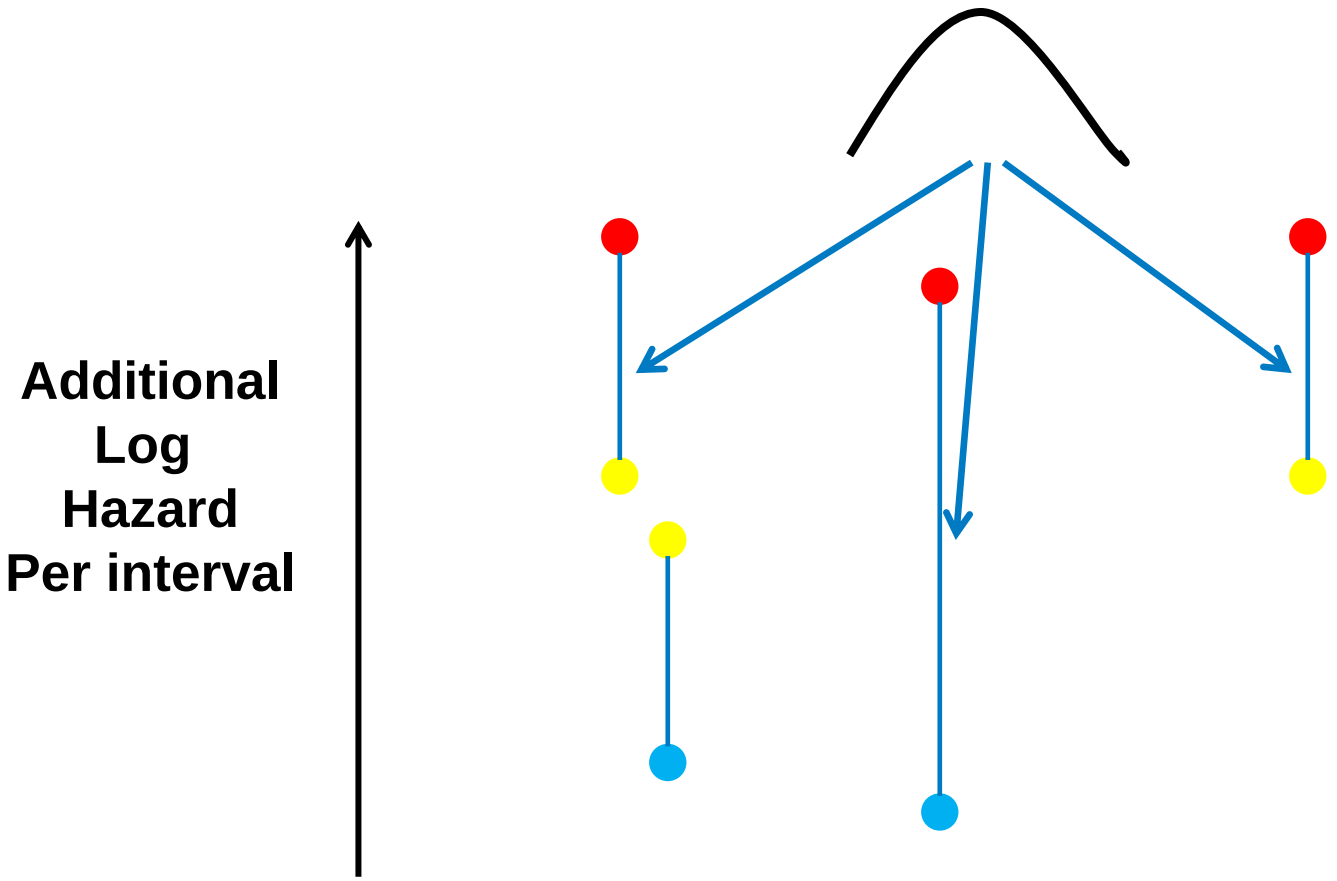
Absolute log hazard model (ALH)



Random Walk Baseline
RW-B



Random Treatment Effect over Time R-LHR(t)



Random Walk Treatment Effect RW-LHR(t)

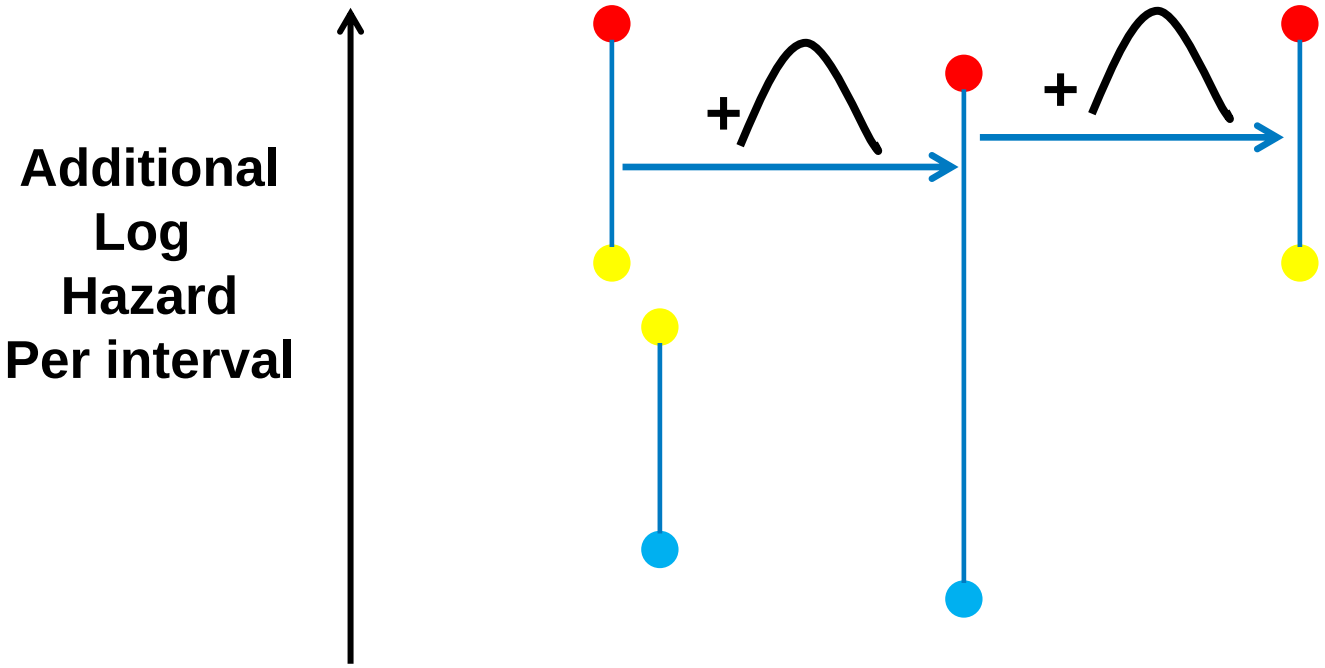


Table II. Global goodness of fit measures on 10 models.

Model	Main features	$\overline{D}$	pD	DIC
1. ALH Hom-V	$\theta_{jk} \sim \mathrm{N}(t_k, \sigma^2)$	410	59	469
2. ALH Het-V	$\theta_{jk} \sim \mathrm{N}(t_k, \sigma_k^2)$	398	60	458
3. F-B, F-LHR	$\theta_{jku} = \mu_j + v_u I(u \geqslant 2) + X_{jk}\Delta_{bk}$	324	50	374
4. F-B, R-LHR, Hom-V	$\theta_{jku} = \mu_j + v_u I(u \geqslant 2) + X_{jk}\delta_{jbk}$ $\delta_{jbk} \sim \mathrm{N}(d_k - d_b, v^2)$	271	67	338
5. F-B, R-LHR(t), Hom-V	$\theta_{jku} = \mu_j + v_u I(u \geqslant 2) + X_{jk}\delta_{jbku}$ $\delta_{jbku} \sim \mathrm{N}(d_k - d_b, v^2)$	233	82	315
6. M-B, F-LHR	$\theta_{jku} = \mu_j + (v_u + \phi_{ju})I(u \geqslant 2) + X_{jk}\Delta_{bk}$	223	69	292
7. ME-B, R-LHR, Hom-V	$\theta_{jku} = \mu_j + (v_u + \varphi_{ju})I(u \geqslant 2) + X_{jk}\delta_{jbk}$ $\delta_{jbk} \sim \mathrm{N}(d_k - d_b, v^2)$	172	87	259
8. ME-B, R-LHR, Het-V	$\theta_{jku} = \mu_j + (v_u + \varphi_{ju})I(u \geqslant 2) + X_{jk}\delta_{jbk}$ $\delta_{jbk} \sim \mathrm{N}(d_k - d_b, v_{bk}^2)$	165	90	255
9. RW-B, R-LHR, Het-V	$\theta_{jku} = \mu_{ju} + X_{jk}\delta_{jbk}$ $\mu_{ju} \sim \mathrm{N}(\mu_{j,u-1}, \tau_{\mathrm{RW}}^2), u \geqslant 2$ $\delta_{jbku} \sim \mathrm{N}(d_k - d_b, v_{bk}^2)$	165	92	257
10. ME-B, RW-LHR(t), Het-V	$\theta_{jku} = \mu_j + (v_u + \varphi_{ju})I(u \geqslant 2) + X_{jk}\delta_{jbku}$ $\delta_{jbk1} \sim \mathrm{N}(d_k - d_b, v_{bk}^2)$ $\delta_{jbku} \sim \mathrm{N}(\delta_{jk,u-1}, \tau_{\mathrm{RW}}^2), u \geqslant 2$	162	92	254

Note: ALH, absolute log hazard; F-LHR, R-LHR, fixed, random log hazard ratio; Hom-V, Het-V, homogeneous, heterogeneous between-trials variance; (t), time-dependent random effects; F-B, ME-B, fixed, mixed effect baselines; RW, random walk.

Results

Table III. Model 10, mixed effects baseline, with random-walk LHR(t) and heterogeneous variance.

Treatment	Average log hazard ratio d_k (95 per cent CI)	Pr (k is best)	Average reduction in unhealed days (95 per cent CI)
1. Placebo	0 (reference)	0.000	0 (reference)
2. PA	1.499 (0.601, 2.415)	0.018	19.4 (6.6, 35.9)
3. H_2RA	1.312 (0.667, 2.159)	0.000	15.4 (8.4, 25.2)
4. H_2RA Double	1.151 (0.185, 2.157)	0.001	13.2 (1.7, 27.4)
5. PPI	2.358 (1.666, 3.215)	0.307	38.0 (27.8, 49.8)
6. PPI Double	2.421 (1.727, 3.272)	0.674	39.5 (28.8, 51.1)

Pharmacoeconomics 2010; 28 (10): 935-94
1170-7690/10/0010-0935/$49.95/

© 2010 Adis Data Information BV. All rights reserved

Comparative Effectiveness Without Head-to-Head Trials

A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept

James E. Signorovitch,[1] *Eric Q. Wu,*[1] *Andrew P. Yu,*[1] *Charles M. Gerrits,*[2] *Evan Kantor,*[1] *Yanjun Bao,*[2] *Shiraz R. Gupta*[2] and *Parvez M. Mulani*[2]

1 Analysis Group Inc., Boston, Massachusetts, USA
2 Abbott Laboratories, Abbott Park, Illinois, USA

Abstract

The absence of head-to-head trials is a common challenge in comparative effectiveness research and health technology assessment. Indirect cross-trial treatment comparisons are possible, but can be biased by cross-trial differences in patient characteristics. Using only published aggregate data, adjustment for such biases may be impossible. Although individual patient data (IPD) would permit adjustment, they are rarely available for all trials. However, many researchers have the opportunity to access IPD for trials of one treatment, a new drug for example, but only aggregate data for trials of comparator treatments. We propose a method that leverages all available data in this setting by adjusting average patient characteristics in trials with IPD to match those reported for trials without IPD. Treatment outcomes, including continuous, categorical and censored time-to-event outcomes, can then be compared across balanced trial populations.

Confounding by Trial Heterogeneity

Table I. An hypothetical adjusted indirect comparison with bias due to differences in patient baseline characteristics

Outcome	Trial of A vs C		Trial of B vs C		A vs B[a]
	drug A	drug C	drug B	drug C	
Response rates (%) aggregated by arm[b]					
All patients	42.5	12.5	57.5	17.5	–10
Response rates (%) stratified by baseline severity using IPD					
Severe	30	10	20	10	10
Non-severe	80	20	70	20	10

a Difference-in-difference of response rates: (A – C) – (B – C).

b Assuming 75% severe (25% non-severe) in the trial of A vs C and 25% severe (75% non-severe) in the trial of B vs C.

IPD = individual patient data.

Adjustment based on IPD

- Individual Patient Data (IPD) is available for one trial and but not another (aggregate data only)
- The trials are ‘matched’ by re-weighting patients in the IPD trial by their odds of being enrolled in the trial without IPD
- Akin to the use of propensity scores in observational research

Results

Table III. Response rates before and after matching

Week 12 PASI response	REVEAL/528 pre-match		REVEAL/528 post-match		Leonardi et al.[28]		Matching-adjusted indirect comparison (ADA vs ETN[b])
	ADA (n = 678)	PL (n = 347)	ADA (n = 388)[a]	PL (n = 203)[a]	ETN (n = 164)	PL (n = 166)	
≥75%	67.0	4.6	66.5	3.5	49.4	3.6	17.2*
≥90%	37.0	1.4	37.1	0.9	22.0	0.6	14.8*

a Computed from the effective total sample size and proportion in each arm after re-weighting.

b Difference-in-difference of response rates: (ADA – PL in post-match REVEAL/528) – (ETN – PL in Leonardi et al.[28]).

ADA = adalimumab; **ETN** = etanercept; **PASI** = Psoriasis Area and Severity Index; **PL** = placebo; * $p < 0.001$.

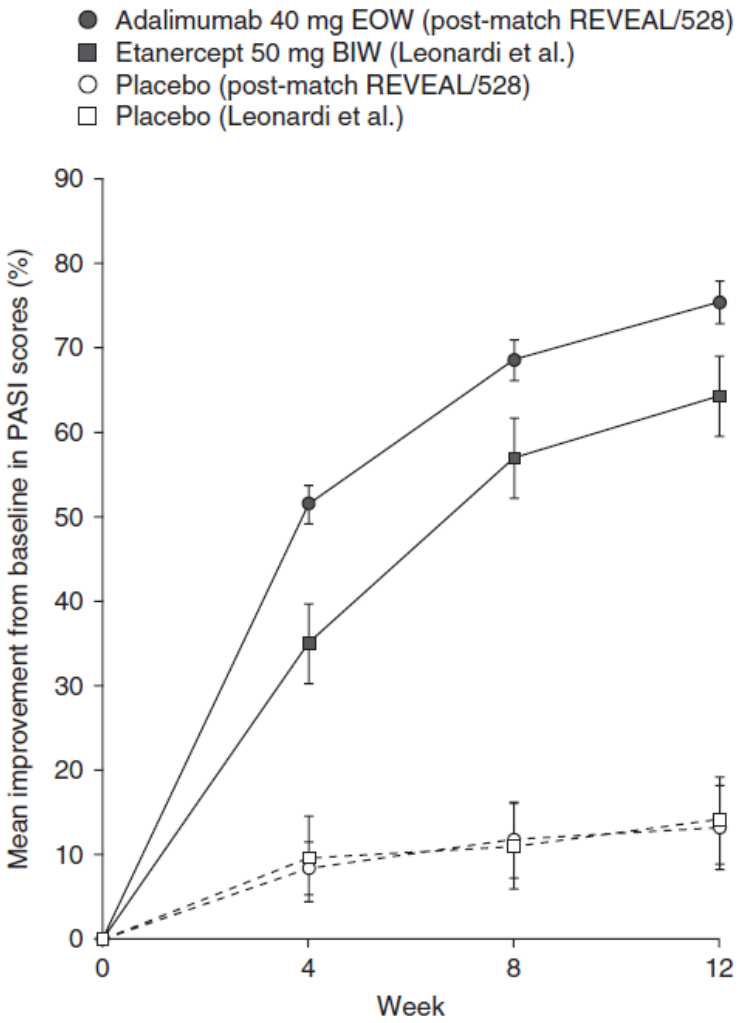


Fig. 1. Matching-adjusted mean percentage improvements in Psoriasis Area and Severity Index (PASI) scores for etanercept, adalimumab and placebo; data obtained from REVEAL,[27] M02-528[30] and Leonardi et al.[28] **BIW** = twice weekly; **EOW** = every other week.

Discussion

- All examples of adjustment for heterogeneity
- More data = more options for adjustment
- Individual patient level data is useful
- Confidence in the usefulness of the consistency constraint is essentially an empirical question

APPLICATIONS OF NMA

NMA IN PHASE 2

See COMMENTARY page 766 and ARTICLE page 820

A Dose–Response Meta-Analysis for Quantifying Relative Efficacy of Biologics in Rheumatoid Arthritis

JW Mandema[1], DH Salinger[2], SW Baumgartner[3,4] and MA Gibbs[2]

We present a dose–response meta-analysis to quantify relative efficacy of biologic disease-modifying antirheumatic drugs (DMARDs) in patients with rheumatoid arthritis (RA). There is a strong rationale for this analysis because, although multiple biologics are available, information on head-to-head comparisons is limited. Data on the percentage of patients attaining American College of Rheumatology (ACR) 20, 50, and 70 responses were extracted from 50 randomized controlled trials representing 21,500 patients, five mechanisms of action, and nine biologics. The analysis showed that all tumor necrosis factor inhibitors (anti-TNFs) share the same dose–response relationship for ACR 20, 50, and 70, differing only in potency. Yet there are significant differences in efficacy among the anti-TNFs due to differences in the clinical dose ranges available. At the suggested starting dose, golimumab was the least efficacious, followed by infliximab, adalimumab, etanercept, and certolizumab. Significant differences in the dose–response relationship were found between anti-TNFs and other biologics, resulting in differences in efficacy and differential impact of dose titration.

Over the past 15 years, several biologic disease-modifying antirheumatic drugs (DMARDs) have become available for patients with rheumatoid arthritis (RA) who had an inadequate response to traditional nonbiologic DMARDs such as methotrexate (MTX), sulfazalazine, leflunomide, and antimalarials.

- Is there a difference in relative efficacy across different end points of disease activity such as the American College of Rheumatology (ACR) 20, 50, and 70 response criteria?
- Is the efficacy of anti-TNFs different in patients with an inadequate response to MTX as compared with those who

Purpose and methods

- Compare the dose–response relationship for the efficacy end points ACR 20, 50, and 70 for the clinically available biologics in adult patients with RA
- A regression method based on dose–response relationships to account for differences in efficacy as a function of dose
  - increases the precision of the estimated treatment effect at a particular dose
  - differences in treatment effect due to differences in patient populations can be quantified through parameters of the dose–response relationship

Clinical evidence

Table 1 Summary of available information for each drug included in the analysis

Drug	Trials	Patients	Dose range	Trials failed DMARD/MTX/anti-TNF	Age (years)	Disease duration (years)
Placebo						
No DMARD	7	569		0/7/0	52 (48 to 55)	9 (3 to 12)
MTX or other DMARD	33	4,408	MTX 16 (8 to 24) mg qw	1/27/5	53 (49 to 57)	9 (2 to 13)
MTX	8	1,726	18 (15 to 20) mg qw	8/0/0	52 (49 to 54)	1 (0.5 to 7)
Abatacept	6	1,242	10 (0.5 to 10) mg/kg q4w	0/4/2	52 (46 to 56)	9 (3 to 13)
Adalimumab	6	2,136	40 (20 to 160) mg q2w	1/5/0	54 (52 to 57)	11 (0.7 to 13)
Anakinra	3	949	75 (3 to 162) mg/day	0/3/0	53 (49 to 56)	7 (4 to 11)
Certolizumab	4	1,512	200 (200 to 400) mg q2w	0/4/0	52 (51 to 53)	6 (6 to 9)
Etanercept	11	2,493	25 (0.5 to 50) mg biw	4/7/0	52 (48 to 55)	9 (0.7 to 15)
Golimumab	4	1,231	100 (50 to 200) mg q4w	1/2/1	52 (48 to 58)	7 (1 to 9)
Infliximab	7	2,179	6 (2 to 20) mg/kg week 0, 2, 6; q8w	2/5/0	52 (47 to 59)	8 (0.4 to 14)
Rituximab	3	707	1,000 (500 to 1,000) mg week 0, 2	0/2/1	52 (51 to 54)	11 (9 to 12)
Tocilizumab	7	2,377	8 (2 to 8) mg/kg q4w	1/5/1	51 (49 to 54)	9 (2 to 13)
Total	50	21,529		9/36/5	52 (46 to 59)	9 (0.4 to 15)

For continuous variables such as dose, age, and disease duration, the median (range) of values across trials are shown.

biw, twice a week; DMARD, disease-modifying anti-rheumatic drug; MTX, methotrexate; q2w, every 2 weeks; q4w, every 4 weeks; q8w, every 8 weeks; qw, once weekly; anti-TNF, anti–tumor necrosis factor.

Network diagram

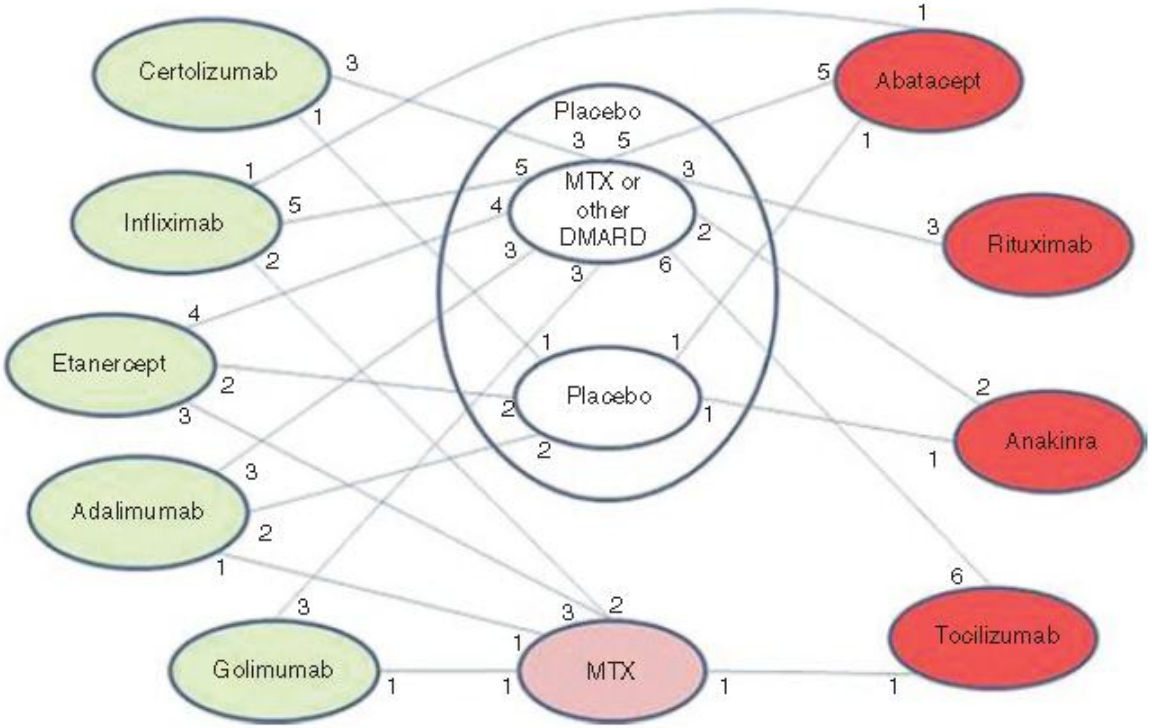


Figure 1 Network diagram of available direct comparisons. The numbers at the lines indicate the number of trials for each direct comparison. DMARD, disease-modifying antirheumatic drug; MTX, methotrexate.

Dose-response relationships - Emax

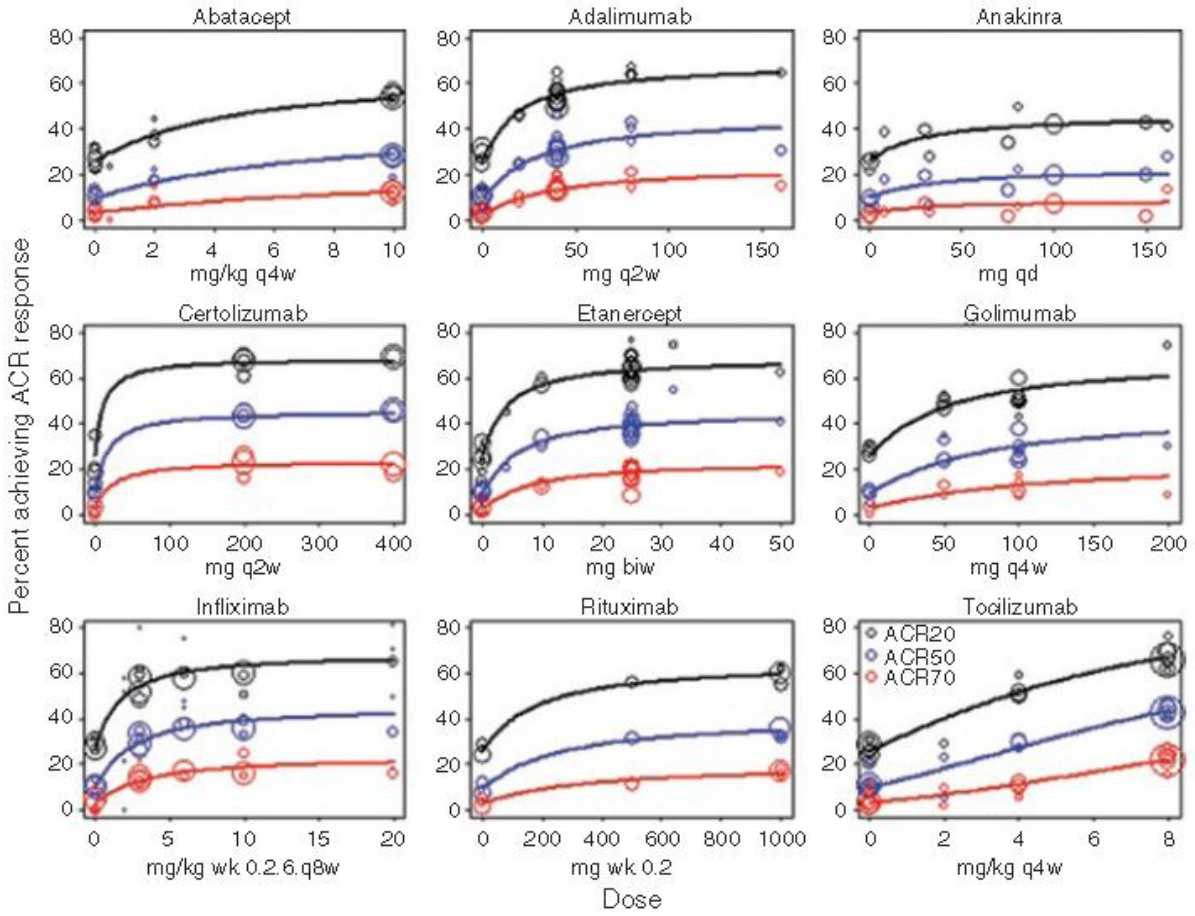


Figure 2 Estimated vs. observed dose–response relationship for American College of Rheumatology (ACR) 20, 50, and 70 responses in patients with rheumatoid arthritis. The symbols represent the observed response incidence in an arm of a trial after adjusting for the between-trial random difference in placebo response and are color-coded by end point. The size of the symbol is proportional to the precision; a larger symbol indicates a more precise (i.e., based on a larger sample size) observation. The lines are the predicted dose–response relationships for ACR 20 (black), ACR 50 (blue), and ACR 70 (red).

Estimated differences from placebo

Table 2 Estimated mean absolute difference in ACR response from placebo (patients on background MTX treatment) for the suggested starting dose of each biologic in a typical patient population with an inadequate response to prior MTX treatment (95% CI)

Drug	Dose	Frequency	ACR 20	ACR 50	ACR 70
Anakinra	100 mg	qd	15.5 (6.9 to 20.4)	9.5 (3.8 to 12.9)	4 (1.5 to 5.7)
Golimumab	50 mg	q4w	21.3 (15.2 to 26.8)	13.9 (9.1 to 18.3)	6.2 (3.8 to 8.5)
Tocilizumab	4 mg/kg	q4w	25.7 (14.3 to 28.2)	17.7 (8.5 to 19.8)	8.2 (3.5 to 9.4)
Abatacept	10 mg/kg	q4w	27.1 (14.5 to 30.3)	18.9 (8.6 to 21.7)	8.9 (3.6 to 10.6)
Infliximab	3 mg/kg	wk 0, 2, 6, q8w	28.5 (23.5 to 32.3)	20.3 (15.5 to 23.7)	9.7 (7 to 11.8)
Adalimumab	40 mg	q2w	29.2 (24.8 to 32.4)	20.9 (16.5 to 23.9)	10 (7.5 to 11.9)
Rituximab	1,000 mg	wk 0, 2	33.2 (12.1 to 37.4)	25 (7 to 29.3)	12.6 (2.8 to 15.5)
Etanercept	25 mg	biw	37.1 (32.6 to 40)	29.3 (24.1 to 32.2)	15.5 (11.9 to 17.6)
Certolizumab	200 mg	q2w	40.4 (36.7 to 43.6)	33.3 (28.5 to 36.9)	18.4 (14.6 to 21.4)

Typical placebo response is 24% for ACR 20, 8.6% for ACR 50, and 2.7% for ACR 70.

ACR, American College of Rheumatology response; biw, twice a week; MTX, methotrexate; q2w, once every 2 weeks; q4w, once every 4 weeks; q8w, once every 8 weeks; qd, daily.

Estimated mean differences from MTX

Table 3 Estimated mean absolute difference in ACR response from MTX for the suggested starting dose of each biologic in a typical MTX-naive patient population (95% CI)

Drug	Dose	Frequency	ACR 20	ACR 50	ACR 70
Golimumab	50 mg	q4w	−9.8 (−16.3 to −3.4)	−9.2 (−14.9 to −3.4)	−5.3 (−8.6 to −2.1)
Tocilizumab	4 mg/kg	q4w	−5.4 (−17 to −1.9)	−5.3 (−15.2 to −1.9)	−3.2 (−8.7 to −1.1)
Infliximab	3 mg/kg	Week 0, 2, 6; q8w	−2.6 (−8.3 to 2.6)	−2.7 (−8.3 to 2.9)	−1.6 (−5.1 to 1.9)
Adalimumab	40 mg	q2w	−2 (−6.6 to 2.1)	−2 (−6.7 to 2.3)	−1.3 (−4.2 to 1.5)
Etanercept	25 mg	biw	5.8 (1.7 to 8.9)	6.5 (1.9 to 10.3)	4.4 (1.3 to 7.3)

ACR, American College of Rheumatology response; biw, twice a week; CI, confidence interval; MTX, methotrexate; q2w, once every 2 weeks; q4w, once every 4 weeks; q8w, once every 8 weeks.

Table 4 Estimated mean absolute difference in ACR response from MTX for the suggested starting dose of each biologic in combination with MTX in a typical MTX-naive patient population (95% CI)

Drug	Dose	Frequency	ACR 20	ACR 50	ACR 70
Golimumab	50 mg	q4w	12.2 (8.8 to 15)	13.1 (9.4 to 17.3)	8.4 (5.7 to 12.2)
Infliximab	3 mg/kg	wk 0, 2, 6;q8w	15.5 (12.6 to 17.6)	17.4 (14 to 20.9)	11.6 (8.8 to 15.4)
Adalimumab	40 mg	q2w	15.7 (12.9 to 17.8)	17.7 (14.4 to 21.2)	11.9 (9.1 to 15.7)
Etanercept	25 mg	biw	19.1 (15.9 to 21.2)	22.5 (18.9 to 26)	15.8 (12.6 to 20)

ACR, American College of Rheumatology response; biw, twice a week; MTX, methotrexate; q2w, once every 2 weeks; q4w, once every 4 weeks; q8w, once every 8 weeks..

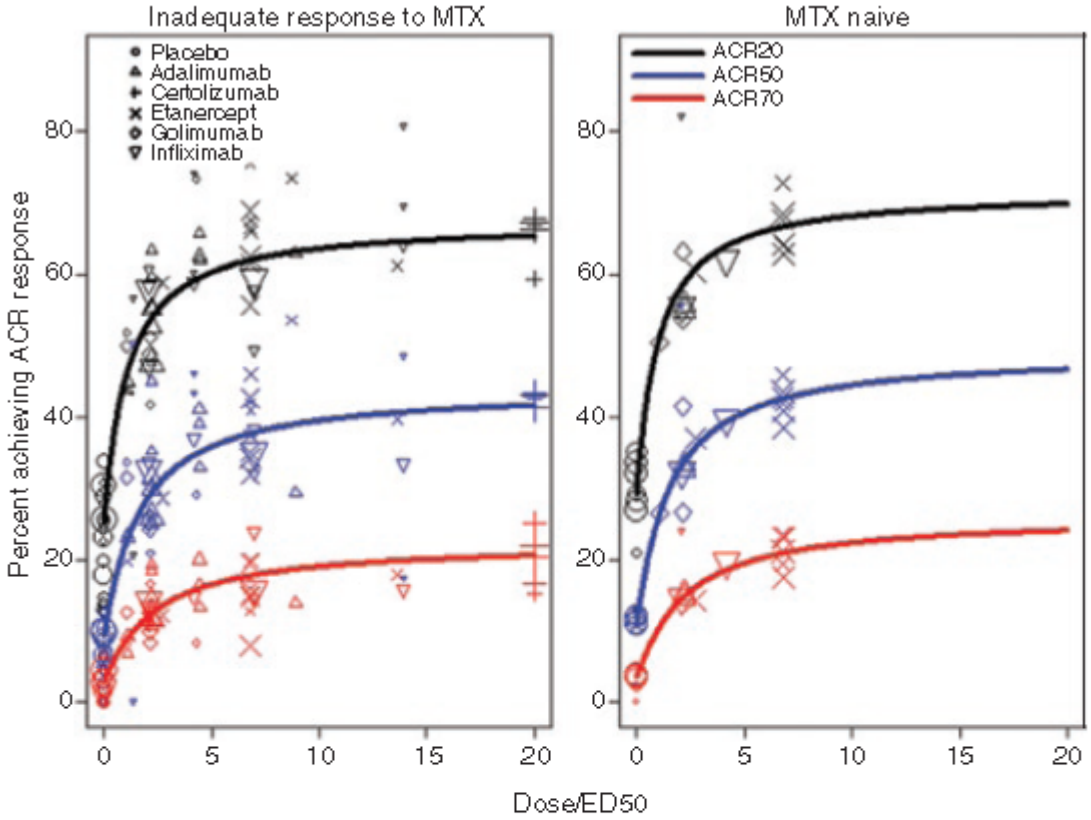


Figure 4 Estimated vs. observed normalized dose–response relationships for tumor necrosis factor inhibitors, stratified by trials in which patients had an inadequate response to prior treatment with methotrexate (MTX) (left panel) and trials in MTX-naive patients (right panel). The symbols represent the observed response incidence in an arm of a trial after adjusting for the between-trial random difference in placebo response and are coded by end point and drug. The dose is expressed as dose/ED_{50} (dose required to achieve 50% of maximum effect). The size of the symbol is proportional to the precision; a larger symbol indicates a more precise (i.e., based on a larger sample size) observation. The lines are the predicted dose–response relationships for American College of Rheumatology (ACR) 20 (black), ACR 50 (blue), and ACR 70 (red) responses.

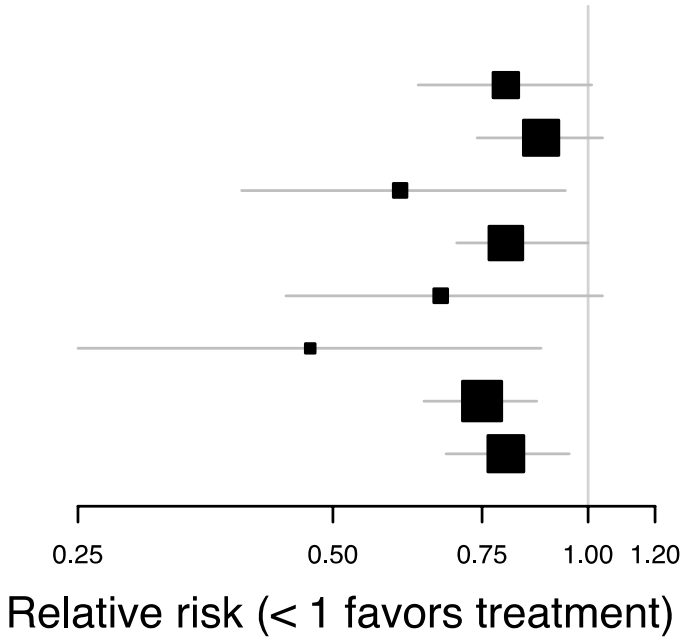
Summary

- The dose–response relationships for ACR 20, 50, and 70 were quantified for the clinically available biologic DMARDs.
- The dose–response-based meta-analysis provided insights into the relative efficacies across the different mechanisms of action and among the five anti-TNFs.
- Head-to head comparative trials are needed to confirm these results.

NMA IN PHASE 2/3

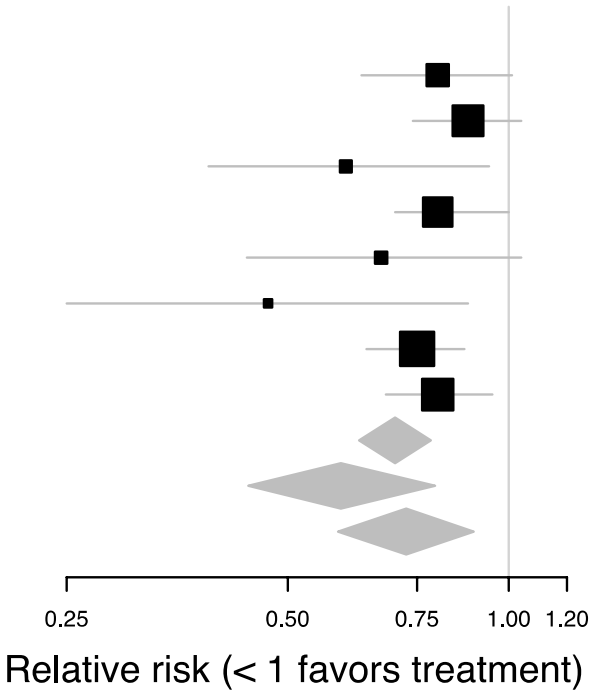
Hypothetical – phase 3 existing data

Treatment	Control
Drug 1	PBO
Drug 1	PBO
Drug 2	PBO
Drug 2	PBO
Drug 3	PBO
Drug 4	PBO
Drug 4	PBO
Drug 5	PBO

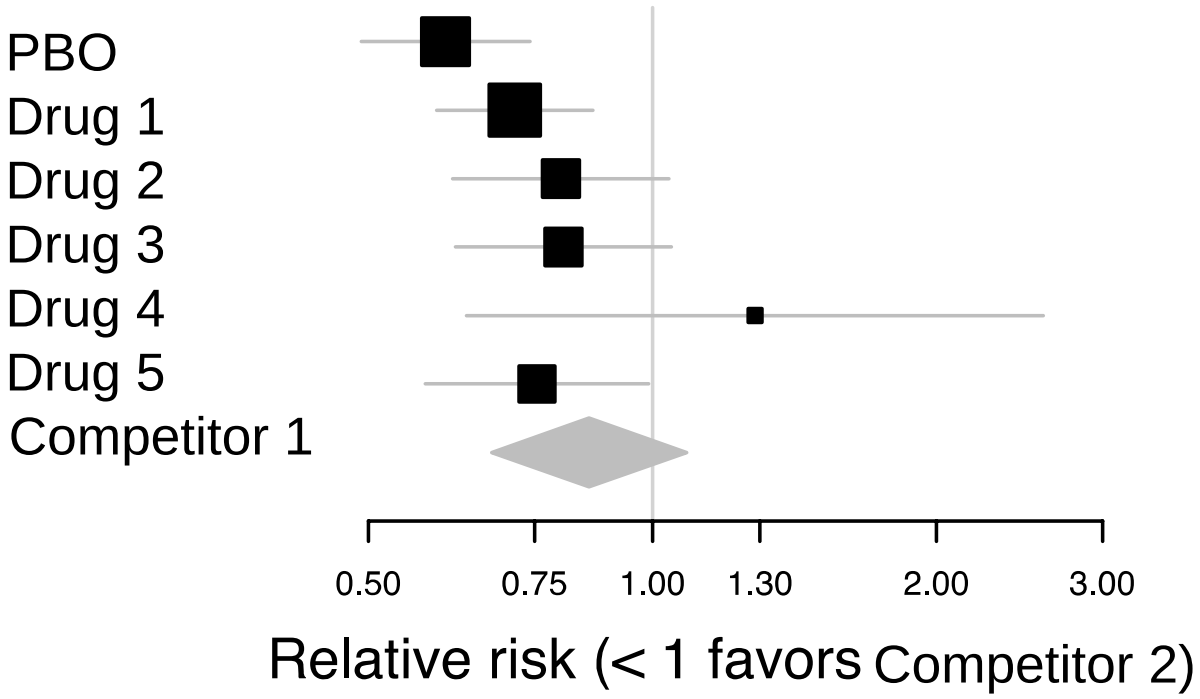


Simulate competitor key trial results

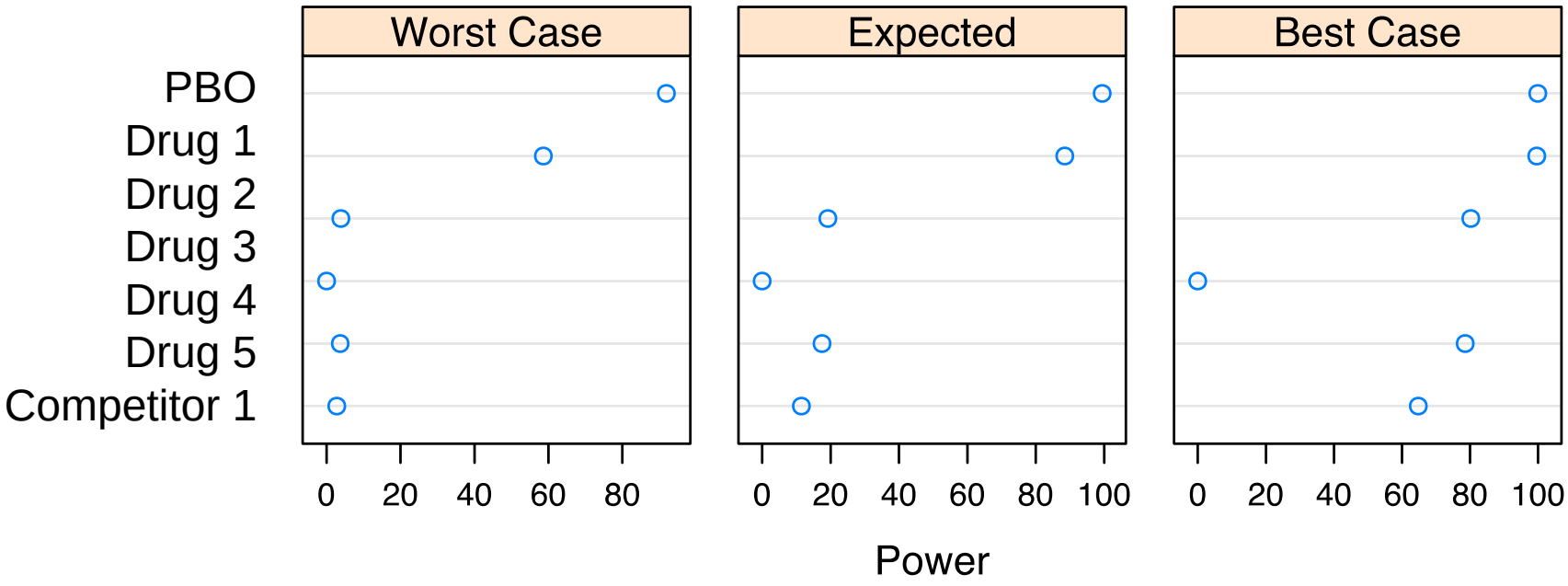
Treatment	Control
Drug 1	PBO
Drug 1	PBO
Drug 2	PBO
Drug 2	PBO
Drug 3	PBO
Drug 4	PBO
Drug 4	PBO
Drug 5	PBO
Competitor 1	PBO
Competitor 2	PBO
Competitor 2	PBO



Simulate indirect comparisons for future competitors



Simulate power for Competitor 2



Summary

- Simulating NMAs in phase 2/3 can improve trial designs
- Preliminary NMAs can enable the impact of competitor data to be considered
- Preliminary NMAs aid the planning for comparative effectiveness activities required for reimbursement

NMA FOR HTA - DENOSUMAB (PROLIA®) OSTEOPOROSIS

Osteoporos Int
DOI 10.1007/s00198-012-2068-9

ORIGINAL ARTICLE

Results of indirect and mixed treatment comparison of fracture efficacy for osteoporosis treatments: a meta-analysis

N. Freemantle · C. Cooper · A. Diez-Perez · M. Gitlin · H. Radcliffe · S. Shepherd · C. Roux

Received: 22 February 2012 / Accepted: 4 June 2012
© International Osteoporosis Foundation and National Osteoporosis Foundation 2012

Abstract
Summary Network meta-analysis techniques (meta-analysis, adjusted indirect comparison, and mixed treatment comparison [MTC]) allow for treatment comparisons in the absence of head-to-head trials. In this study, conditional estimates of relative treatment efficacy derived through these techniques show important differences in the fracture risk reduction profiles of marketed pharmacologic therapies for postmenopausal osteoporosis.

Introduction This study illustrates how network meta-analysis techniques (meta-analysis, adjusted indirect comparison, and MTC) can provide comparisons of the relative efficacy of postmenopausal osteoporosis therapies in the absence of comprehensive head-to-head trials.
Methods Source articles were identified in MEDLINE; EMBASE; Cochrane Central Register of Controlled Trials (CENTRAL) via Wiley Interscience; and Cumulative Index to Nursing and Allied Health Literature (CINAHL) between April 28, 2009 and November 4, 2009. Two reviewers identified English-language articles reporting randomized controlled trials (RCTs) with on-label dosing of marketed osteoporosis agents and fracture endpoints. Trial design, population characteristics, intervention and comparator, fracture outcomes, and adverse events were abstracted for analysis. Primary analyses included data from RCTs with fracture endpoints. Sensitivity analyses also included studies with fractures reported through adverse event reports. Meta-

Electronic supplementary material The online version of this article (doi:10.1007/s00198-012-2068-9) contains supplementary material, which is available to authorized users.

N. Freemantle (✉)
Department of Primary Care and Population Health, University College London,
London NW3 2PF, UK
e-mail: nicholas.freemantle@ucl.ac.uk

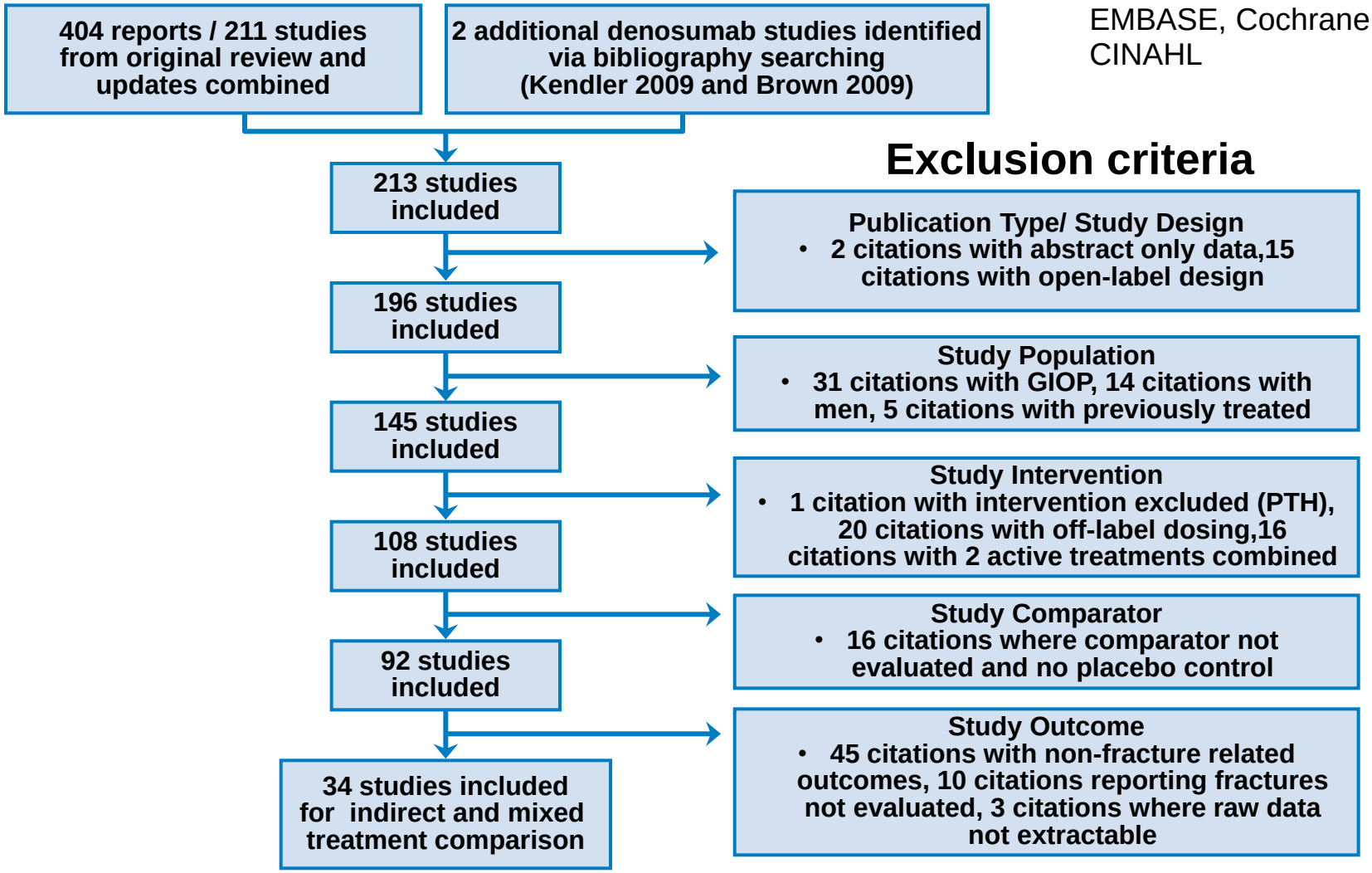
Denosumab (Prolia®) NICE Health Technology Assessment (HTA)

- Initial NICE scoping meeting Jan 2009
- UK HTA core team created May 2009
- Systematic review protocol created Jun 2009
  - Initial search completed
- Research Project Plan created Oct 2009
- Final NICE Scope issued in Nov 2009
  - Final and updated systematic review completed
- HTA submitted Jan 2010
- Preliminary recommendations (ACD) May 2010
- Final guidance (FAD) Oct 2010

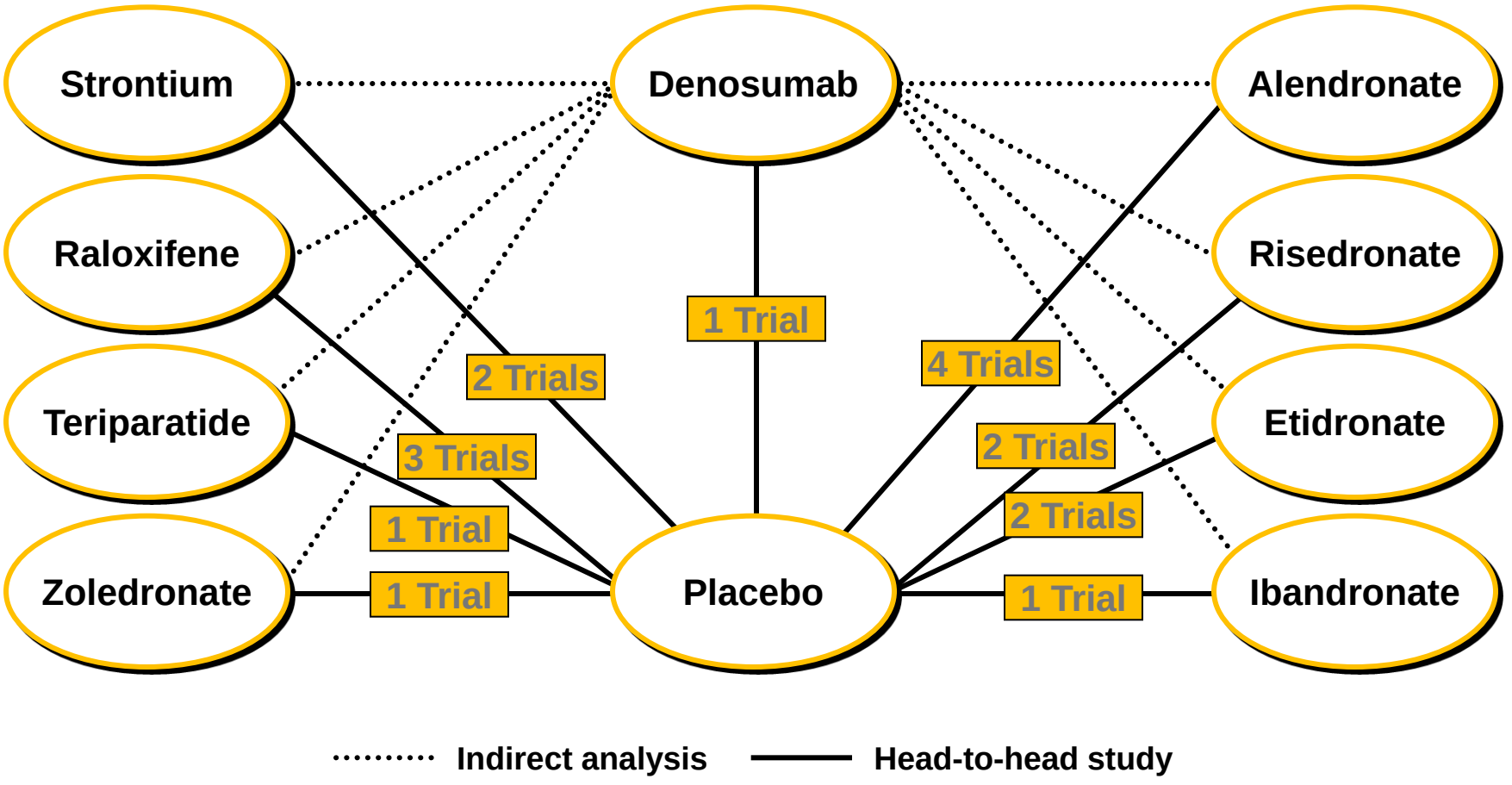
http://guidance.nice.org.uk/TA/Wave20/75

Flow of citations through review process

- 6328 citations initially identified from MEDLINE, EMBASE, Cochrane, CINAHL



Evidence network diagram - new vertebral fracture example



Fracture meta-analysis, adjusted IC & MTC: Comparators

- Comparators
  - Strontium
  - Raloxifene
  - Teriparatide
  - Zoledronate
  - Oral BP
    - Alendronate
    - Risedronate
    - Etidronate
    - Ibandronate

Fracture meta-analysis, adjusted IC & MTC: Endpoints

- Five main fracture types
  - Morphometric vertebral fractures
  - Clinical vertebral fractures
  - Nonvertebral fractures
  - Hip fractures
  - Wrist fractures

  - ‘Other’ fractures was also investigated but no consistent definition was available across studies or publications

Fracture meta-analysis, adjusted IC & MTC: Analysis sets

- Primary analysis – evaluable population for morphometric vertebral fractures, ITT for others
- Sensitivity analysis
  - ITT Population (morphometric vertebral fractures only)
  - All trials, including trials where fractures were captured as adverse events
  - All trials, including trials where fractures were captured as adverse events but excluding trials with additional sources of bias

Fracture meta-analysis, adjusted IC & MTC: Output Requirements

- Comparative efficacy section of the NICE Single Technology Assessment (STA)
  - Meta analysis of fracture data for each comparator relative to a common control (placebo)
  - Adjusted indirect comparison for Dmab vs. comparator
  - Mixed treatment comparison
    - Comparator vs. placebo
    - Dmab vs. comparator
- Input parameters for economic model
  - RR for each comparator vs. placebo
- Subgroup analysis (t-score, age, prev fracture)

Fracture meta-analysis, adjusted IC & MTC: Available data for analysis (morph)

# of trials	Trials		Denosumab	Strontium	Raloxifene	Teriparatide	Ibandronate IV	Zoledronate	Alendronate	Risedronate	Etidronate	Ibandronate Oral	Placebo
1	Cummings_2009	FREEDOM Study	X										X
	Meunier_2004	SOTI Study											
2	Reginster_2008	TROPOS Study		X									X
		MORE Study (Group 1)											
	Ettinger_1999_1	MORE Study											
	Ettinger_1999_2	(Group 2)											
	Lufkin_1998	Lufkin_1998											
4	Morii_2003	Morii_2003			X								X
1	Neer_2001	FPT Study				X							X
1	Black_2007	HORIZON Study						X					X
	Bone_1997	Bone_1997											
	Durson_2001	Durson_2001											
	Black_1996	FIT I Study											
4	Cummings_1998	FIT II Study							X				X
	Harris_1999	VERT NA Study											
2	Reginster_2000	VERT MN Study								X			X
	Herd_1997	Herd_1997											
2	Watts_1990	Watts_1990									X		X
1	Chestnut_2004	BONE Study										X	X

X denotes the treatments compared

Ibandronate oral is 2.5

Indirect Treatment Comparison

- Method
  - Step 1: Perform meta analysis with a common comparator, i.e. placebo
  - Step 2: Approach of Bucher et al adopted for RR

Log RR of indirect comparison of A and B is

$$\ln RR_{AB} = \ln RR_{AC} - \ln RR_{BC}$$

Standard error is

$$SE(\ln RR_{AB}) = \sqrt{[SE(\ln RR_{AC})^2 + SE(\ln RR_{BC})^2]}$$

Bucher HC, Guyatt GH, Griffith LE and Walter SD. The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials. *J Clin Epidemiol.* (1997) 50 (6); 683-691.

Mixed Treatment Comparison

- Method
  - Conducted in Winbugs with standard parameters
  - Based on OR methods in Lu and Ades (2004) and updated for relative risk
  - Method allows for check of heterogeneity of control arms
  - Zero count in one arm acceptable, but not in both arms
  - Non-informative priors used throughout

Mixed Treatment Comparison

- Modelling

Data	Study	Trt	Count	Total
	1	1	4	85
	1	9	6	90
	2	2	78	981
	2	9	145	965
	..			

Model For study 1

$r_i \sim \text{binomial}(p_i, n_i)$

$\text{Ln}(p_i) = \mu_1$ (control)

$\text{Ln}(p_i) = \mu_1 + \text{delta}_1$ (exp)

$\text{delta}_1 \sim \text{norm}([d_1 - d_9], \tau^2)$

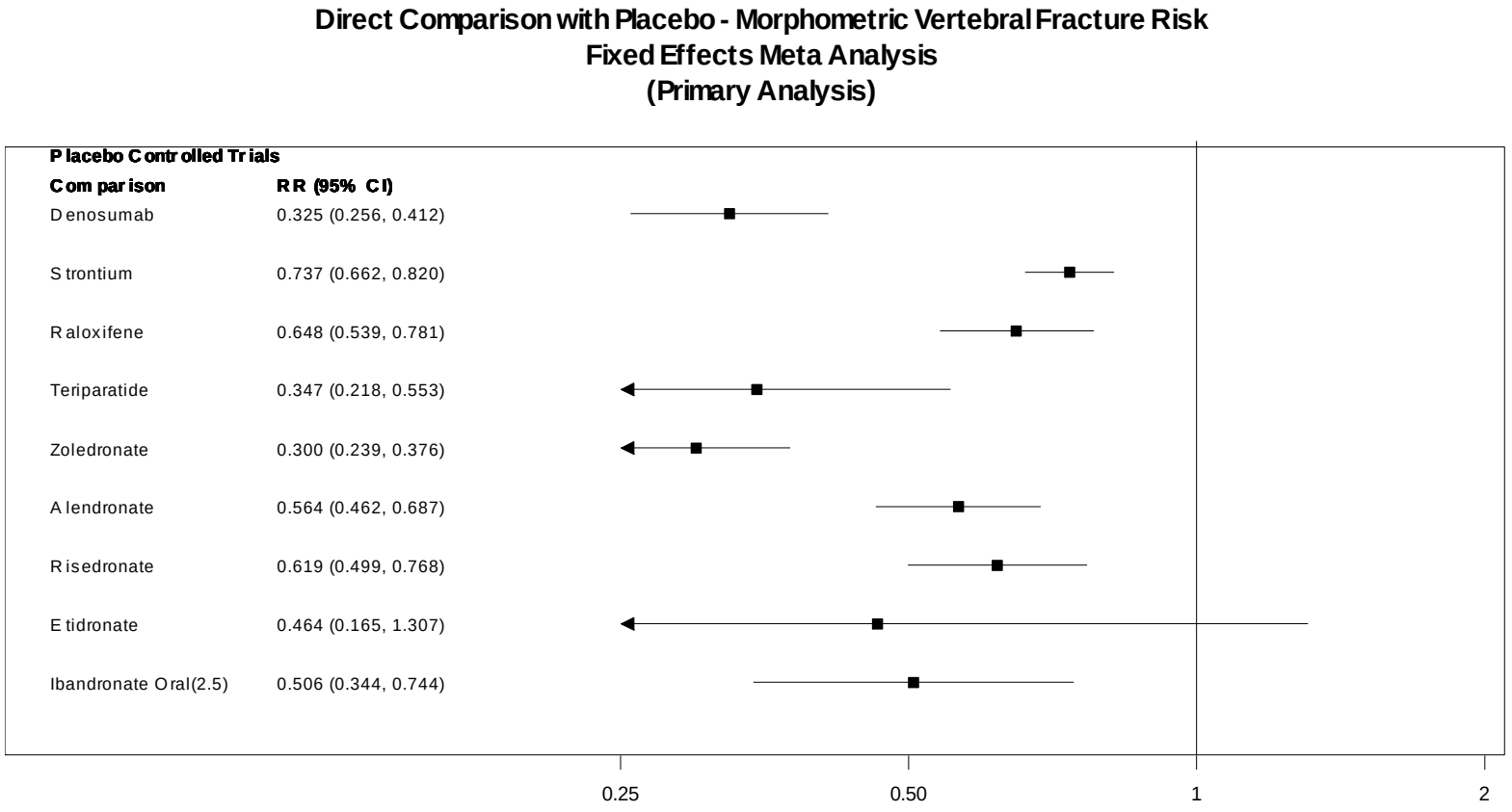
Results Estimates for $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9$

Testing for heterogeneity

- Due to the small number of trials for each comparator, heterogeneity was statistically assessed by the I^2 statistic and is presented in forest plots
- The independent predictors of fracture risk assessed are trial level mean age, proportion of subjects with a prevalent vertebral fracture and mean BMD
- Meta-regression techniques were used to investigate the relationship between the trial level covariates and trial level placebo fracture rate, RD and RR for the primary analysis set
- The estimate and significance level of each covariate are summarised

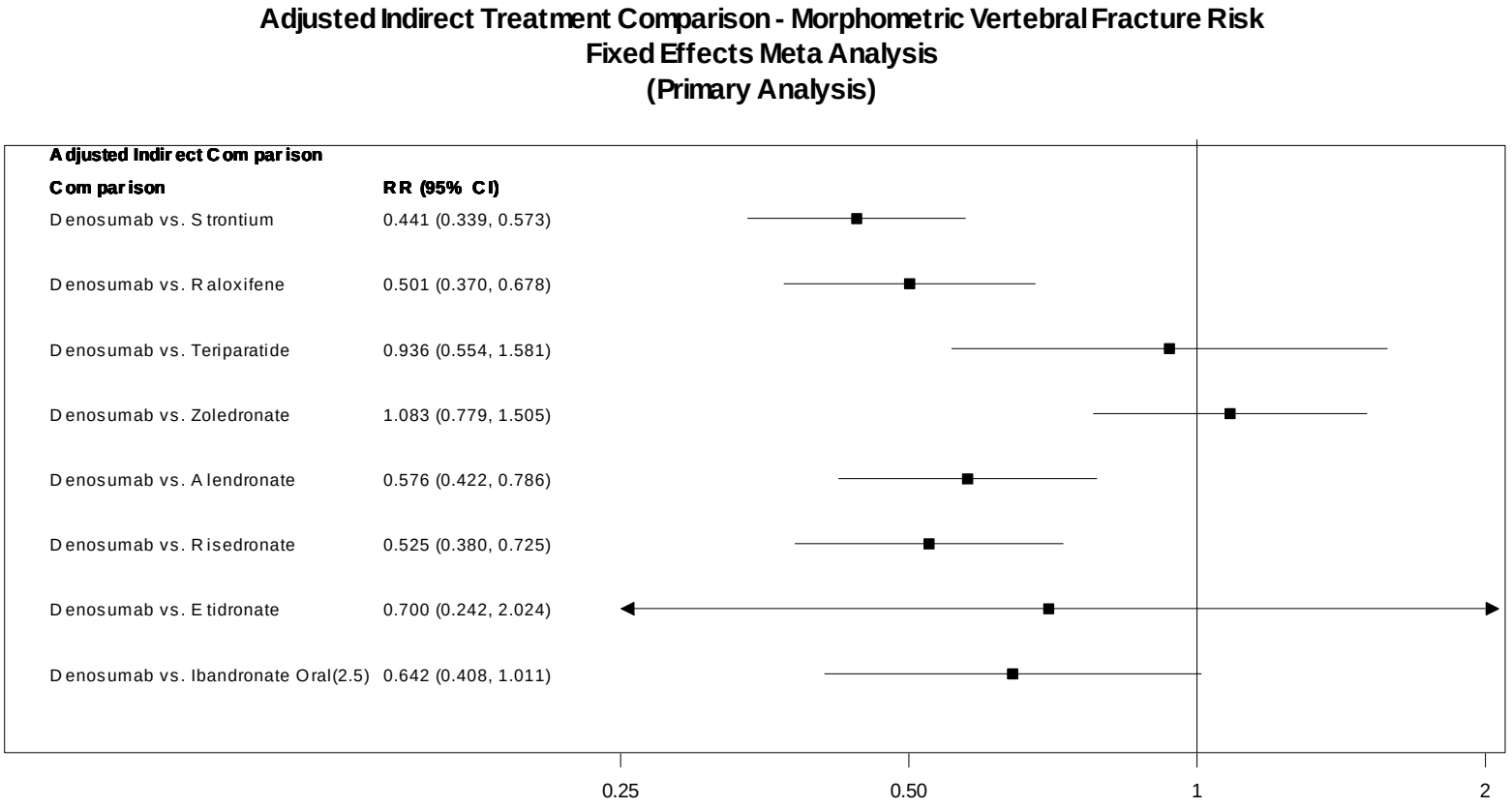
443

Direct evidence

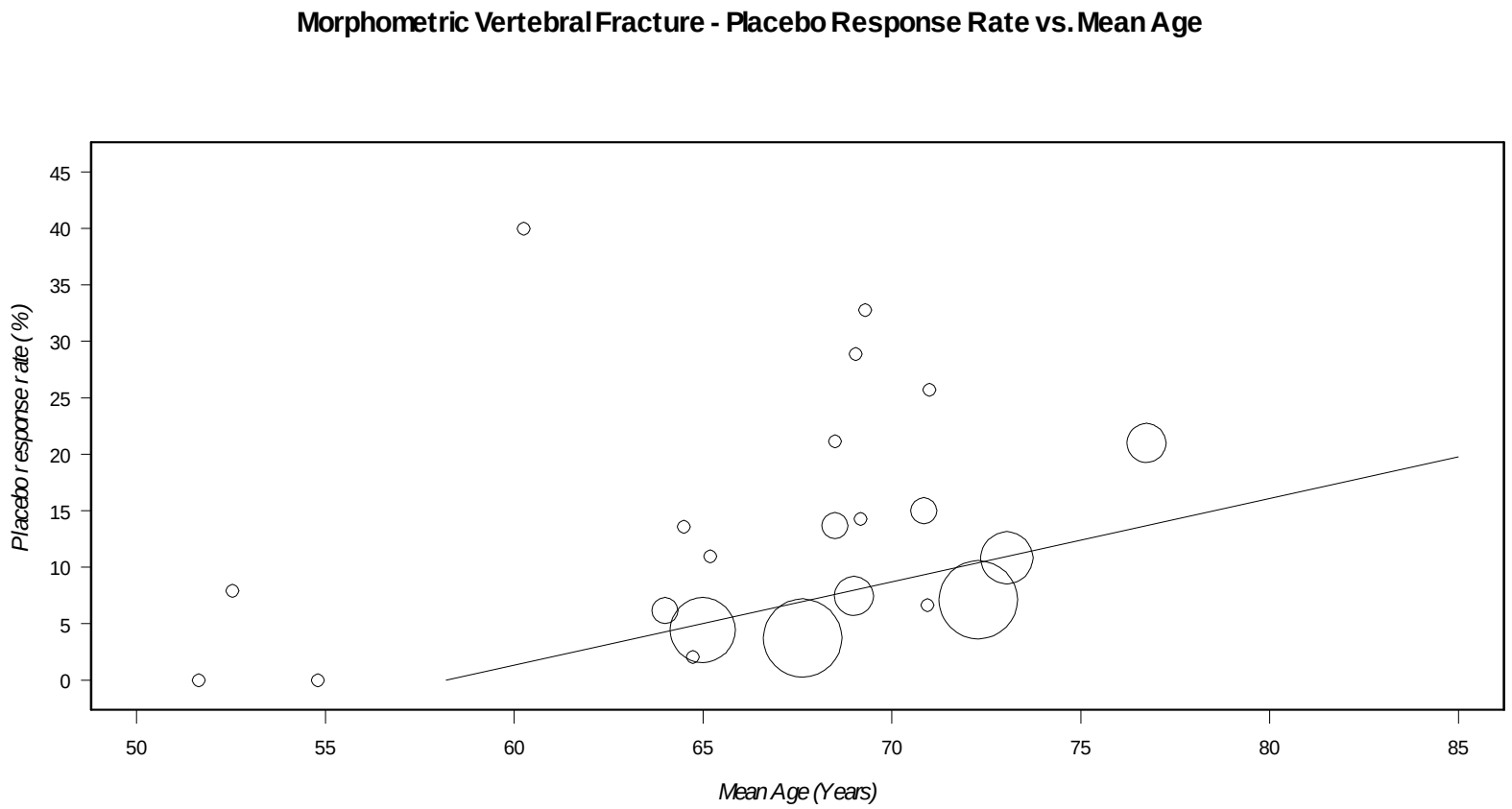


RR = Relative Risk, CI = Confidence Interval
RR < 1 favours comparator

Adjusted indirect evidence

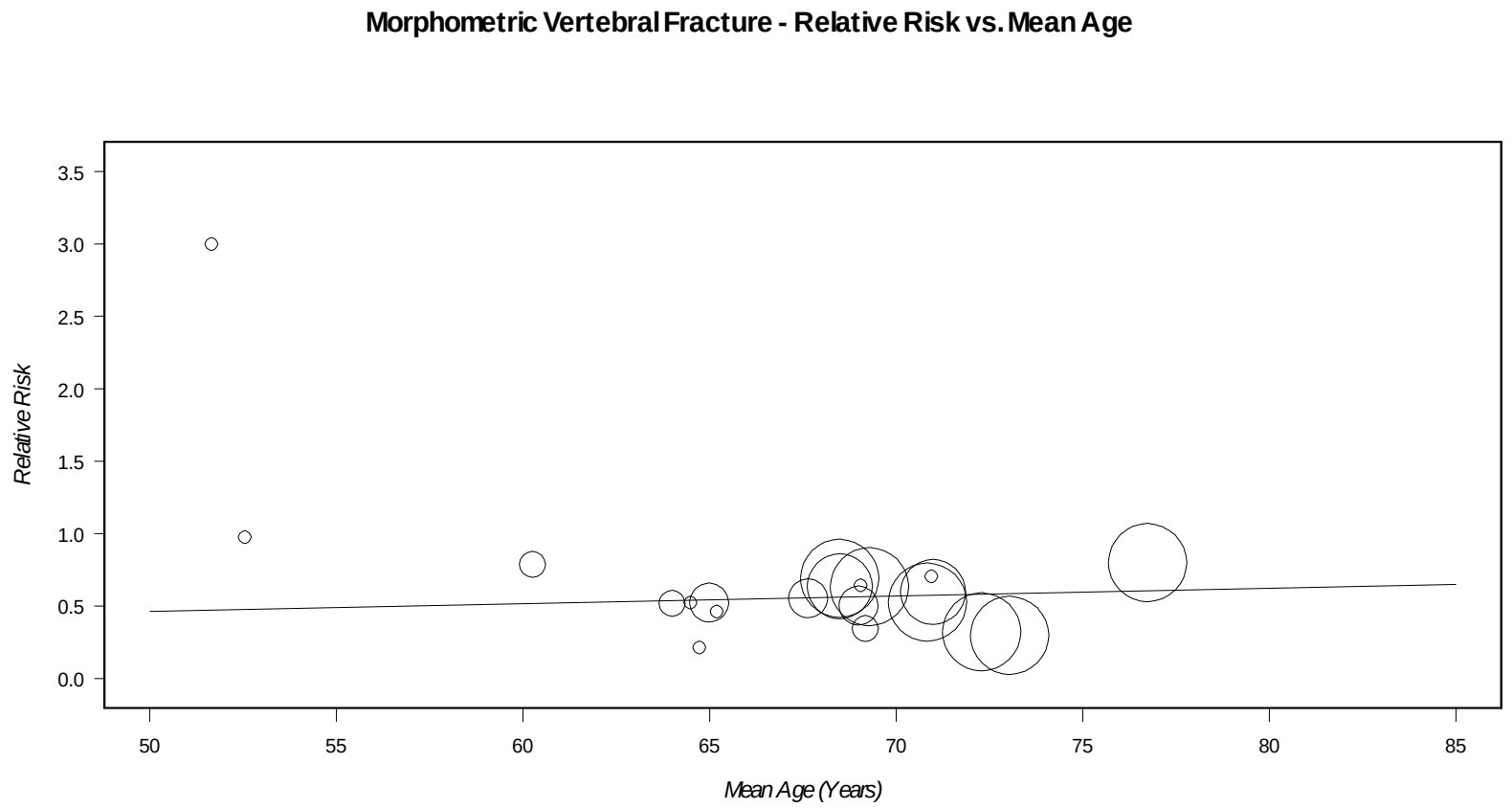


Assessing sources of heterogeneity



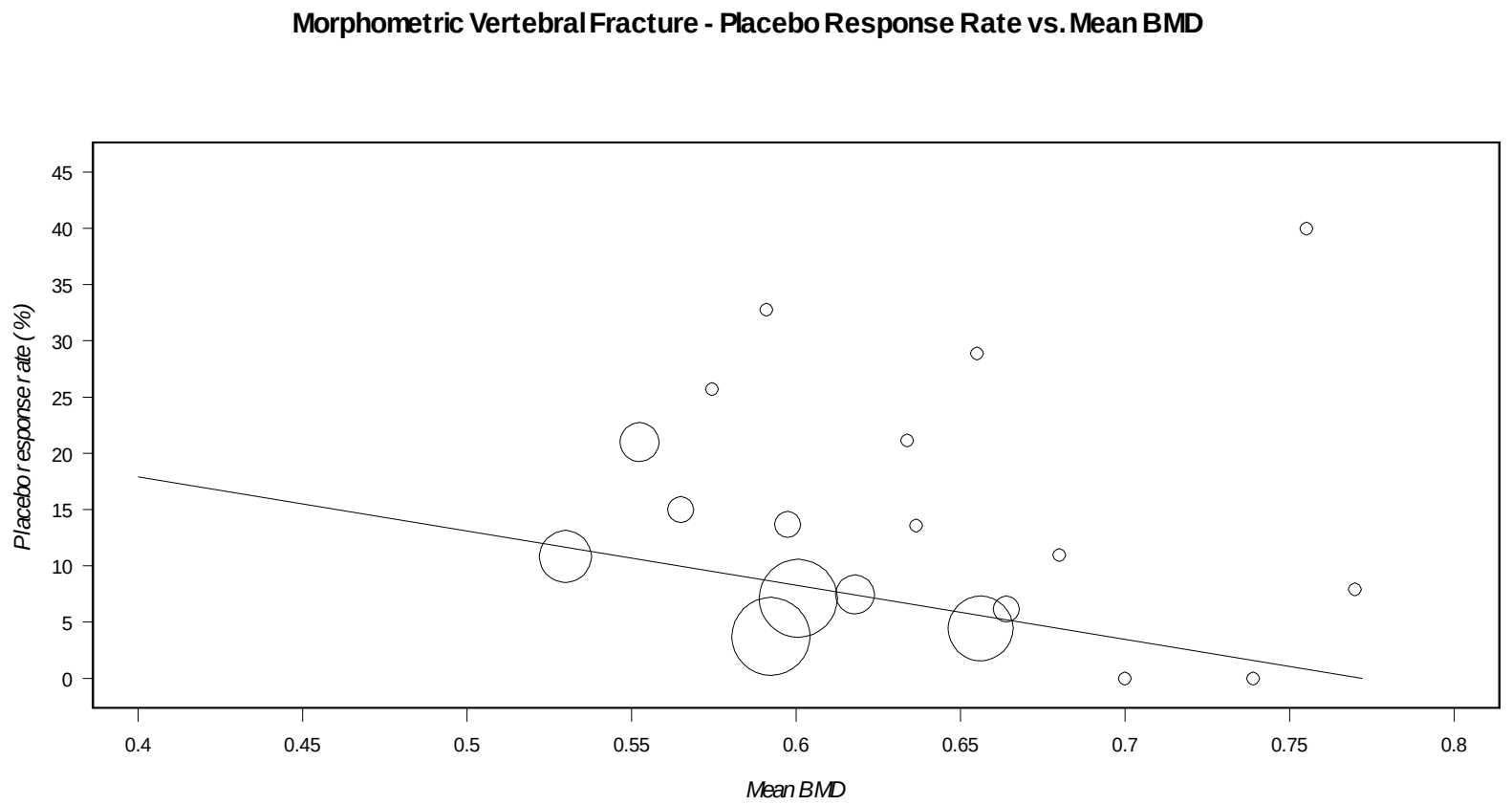
The size of the circles are propotional to the inverse variance
Regression Line is estimated using meta regression

Assessing sources of heterogeneity



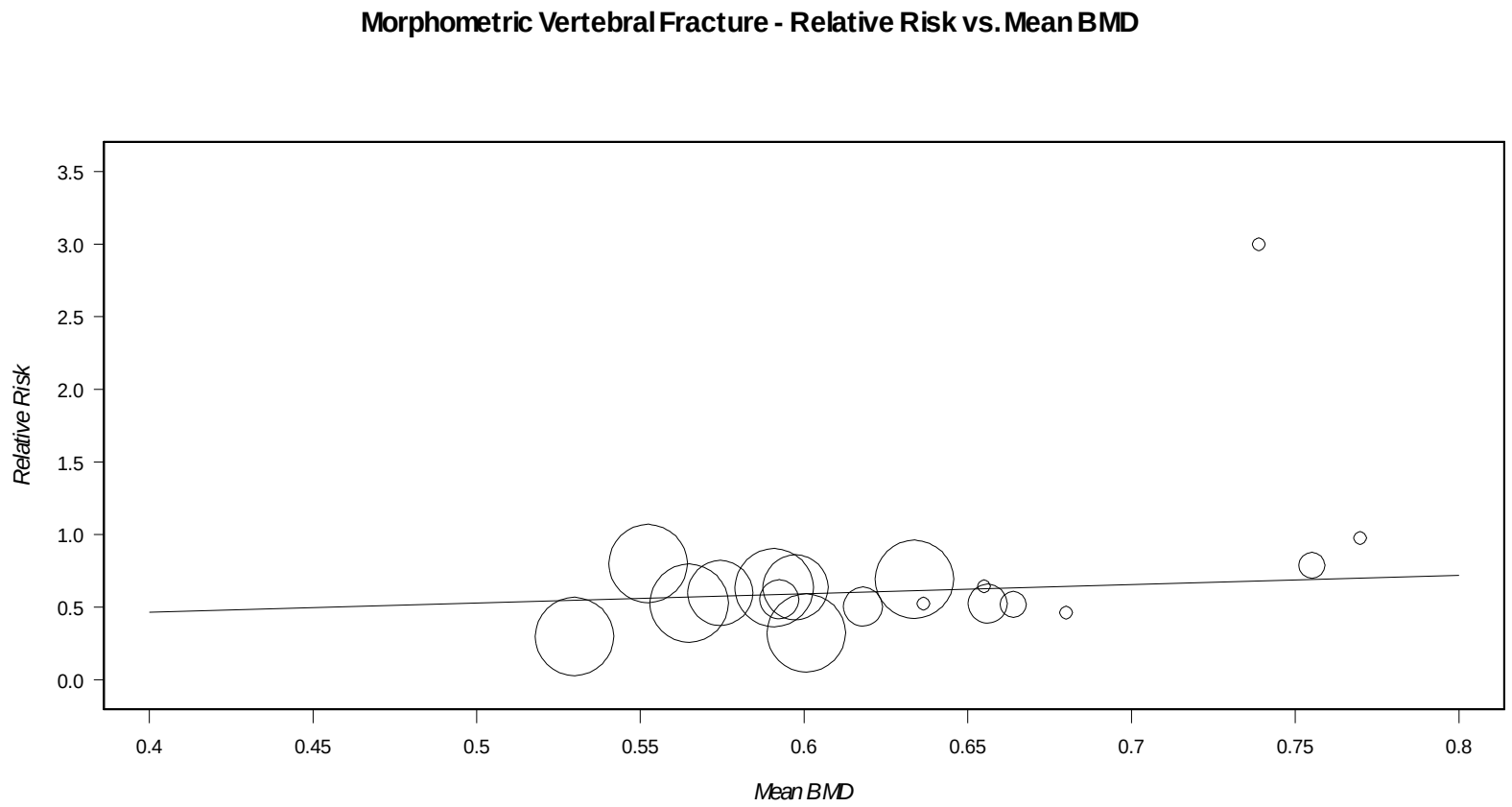
The size of the circles are propotional to the inverse variance
Regression Line is estimated using meta regression

Assessing sources of heterogeneity



The size of the circles are propotional to the inverse variance
Regression Line is estimated using meta regression

Assessing sources of heterogeneity



The size of the circles are propotional to the inverse variance
Regression Line is estimated using meta regression

Summary

- NMAs for reimbursement take time
  - Opportunities to take some work off the critical path
- Additional analyses required
- NMA protocol can accommodate multiple reimbursement agency needs
- Presentation and reporting of NMAs is important
  - Assumptions and technical details should be documented

NMA FOR HEALTHCARE DECISION MAKING

Problem

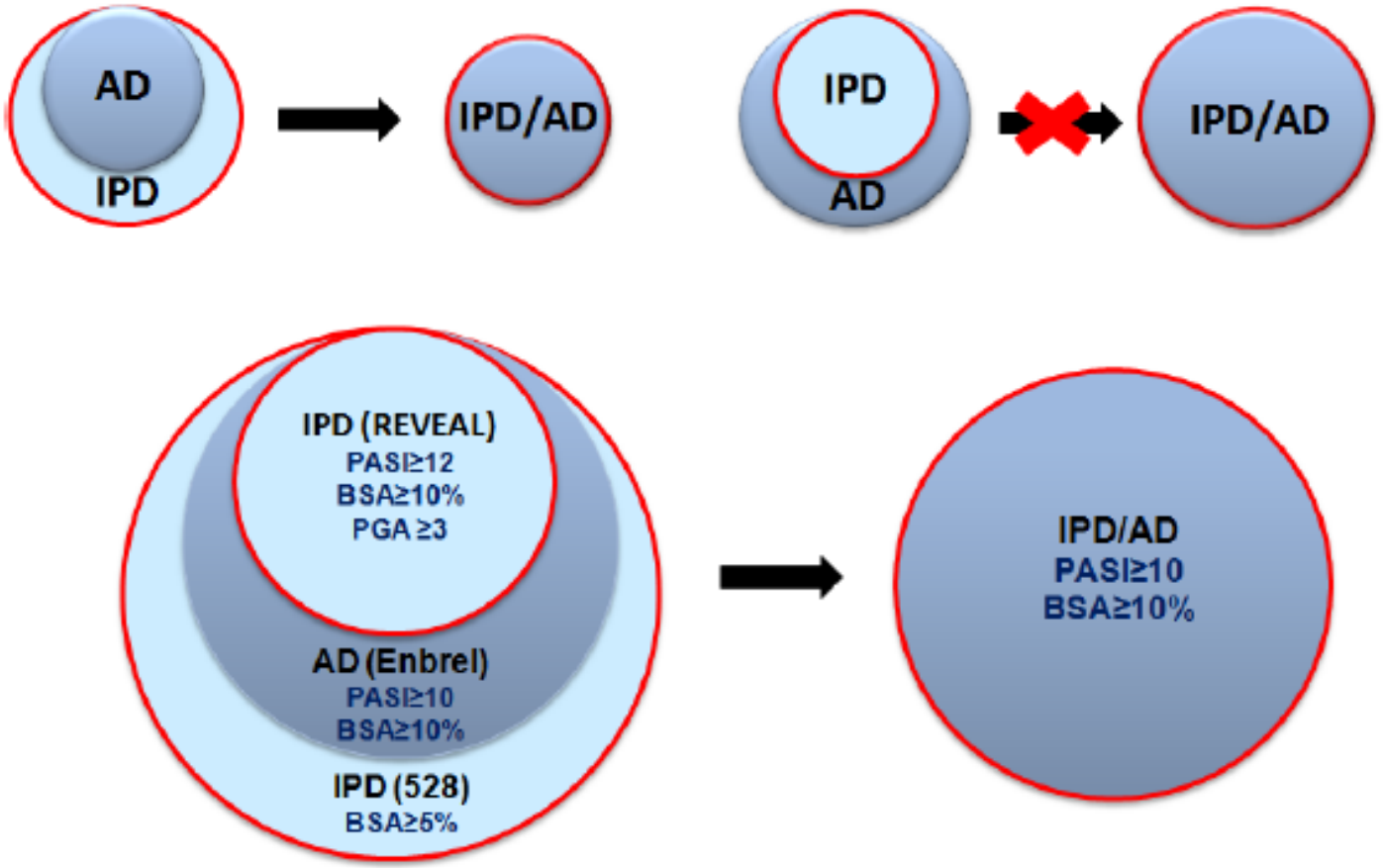
- Individual patient data (IPD) in one trial (or a set of trials) vs. published aggregate data (AD) in another trial (or a set of trials)
- Due to different inclusion/exclusion criteria, background populations may differ resulting in significant differences in crucial covariates/confounders
- How to adjust IPD data with appropriate weights so that the covariates are matched to perform indirect comparisons

Matching Adjusted Indirect Comparison

- Example 1: James E Signorovitch, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics, 28(10):935-945, 2010.

Example-1: Selection of IPD

Match inclusion/exclusion criteria



Example-1 Matching Baseline

Baseline characteristics before and after matching

	REVEAL/528		
Baseline characteristic	Pre-match n=1025	Post-match n=1025	20021639 n=330
Age mean (SD), y	43.9 (13.2)	45.2 (11.6)	45.2 (11.6)
Age ≥65y %	6.4	5.5	5.5
Male %	66.1	64.0	64.0
White %	89.9	88.5	88.5
Duration of psoriasis mean (SD), y	18.3 (12.0)	18.5 (11.5)	18.5 (11.5)
Psoriatic arthritis %	26.9	22.0	22.0
Prior systemic or phototherapy %	62.2*	76.0	76.0
Invoved BSA mean (SD), %	25.9 (15.0)*	29.3 (19.3)	29.3 (19.3)
PASI (SD)	18.9 (6.9)	18.3 (8.4)	18.3 (8.4)
Dermatology life quality index mean (SD)	11.5 (6.7)	12.1 (7.1)	12.1 (7.1)

*$p < 0.05$ for comparisons between pre-match Humia and Enbrel using t-tests for continuous variables and chi-squared tests for categorical variables.

Example-1 Comparison of Outcomes

The matching-adjusted indirect comparison estimator was applied to the comparison of Humira and Enbrel in efficacy outcomes described in Leonardi et al. [2003]

Response rates before and after matching

	REVEAL/528				20021639		$\Delta_1 - \Delta_2$
	Pre-match		Post-match				
Week 12 PASI response	Plb n=347	Hum n=678	Plb n=347	Hum n=678	Plb n=166	Enb n=164	
PASI75, %	4.6	67	3.5	66.5	3.6	49.4	17.2*
PASI90, %	1.4	37	0.9	37.1	0.6	22	14.8*

Δ_1: Hum-Plb ,Δ_2: Enb-Plb. *$p < 0.05$

Summary

- Increasing use of NMA to conduct comparative effectiveness assessments (CEA)
- Need to be able to review and critique a CEA
- Likely will need to replicate analyses

Workshop 4 – NMA opportunities

- Using private collection
  - What opportunities do you have to conduct NMA in your role/company?
  - Is NMA a regular consideration in your drug development activities? If not, what could you do to influence this?
  - How well understood is NMA in your company? Are there opportunities to increase the understanding via educational materials/seminars?
- Write-down an action plan you can take back on following up on NMA opportunities

NMA IN HTA AGENCY METHODOLOGY GUIDELINES

NICE Methods Guide (April 2013)

- Reference case should contain H2H trials
- NMA if required is additional to base case
- NMA will increase uncertainty associated with the lack of direct evidence
- Use best practices for meta-analyses
- Network should contain all interventions/comparators included in the scope
  - Clear methods explaining selecting trials
- Present how direct and indirect evidence compare
- Present results using tabular and graphical displays
  - Present direct evidence separate from NMA

US Agency for Healthcare Research and Quality (AHRQ)

- Study identification from searches of at least 2 databases and supplementary measures
- Greater weight to studies looking at clinical endpoints rather than surrogate endpoints
- All indirect analyses accompanied by sensitivity analysis for robustness assessment
- Random effect methods preferred unless small studies results systematically differ from larger studies results
- Effect measure preference
  - Binary outcome – OR or RR
  - Continuous outcome – actual or standardised mean difference
  - Time to event outcome – HR (with verification of proportionality assumption in the trials)

Canadian Agency for Drugs and Technologies in Health (CADTH)

- Emphasise need to assess similarity of trials for patient/methodological factors and date of the trial
- Trials included must have high external validity
- Can include Mixed Treatment Comparisons, but preference for IC – need to assess inconsistency between direct and indirect evidence
- Random effects methods can be used
- IC methodology based on OR, using anything else needs elaboration
- If no statistical difference found, suggest calculating the power of the IC to detect a difference

Australian Pharmaceutical Benefits Advisory Committee (PBAC)

- Clear justification for inclusion of trials, and they all must have high external validity
- Can use random effects methods if more than 1 trial has evaluated a pair of treatments
- Meta regression is possible if at least 10 trials have measured the covariate
- Sensitivity analyses assess impact of including any controversial trials
- Effect measure preference
  - Binary outcome – RR
  - Continuous outcome – actual or standardised mean difference
  - Time to event outcome – HR
- Reporting an IC – treatment effect for each RCT, pooled estimate for each paired comparison, and indirect estimate of treatment effect of interest

Australian PBAC Working Group

- Expanded version of PBAC recommendations
- Use multiple measures of treatment effect – for binary outcome
  - Risk difference
  - Number needed to treat
  - Odds ratio
  - Relative risk of harm
  - Relative risk of benefit
- No similar recommendations for categorical, continuous or time to event outcome
- Should choose measure of effect to minimise differences between trials

IWQIG (methods V4)

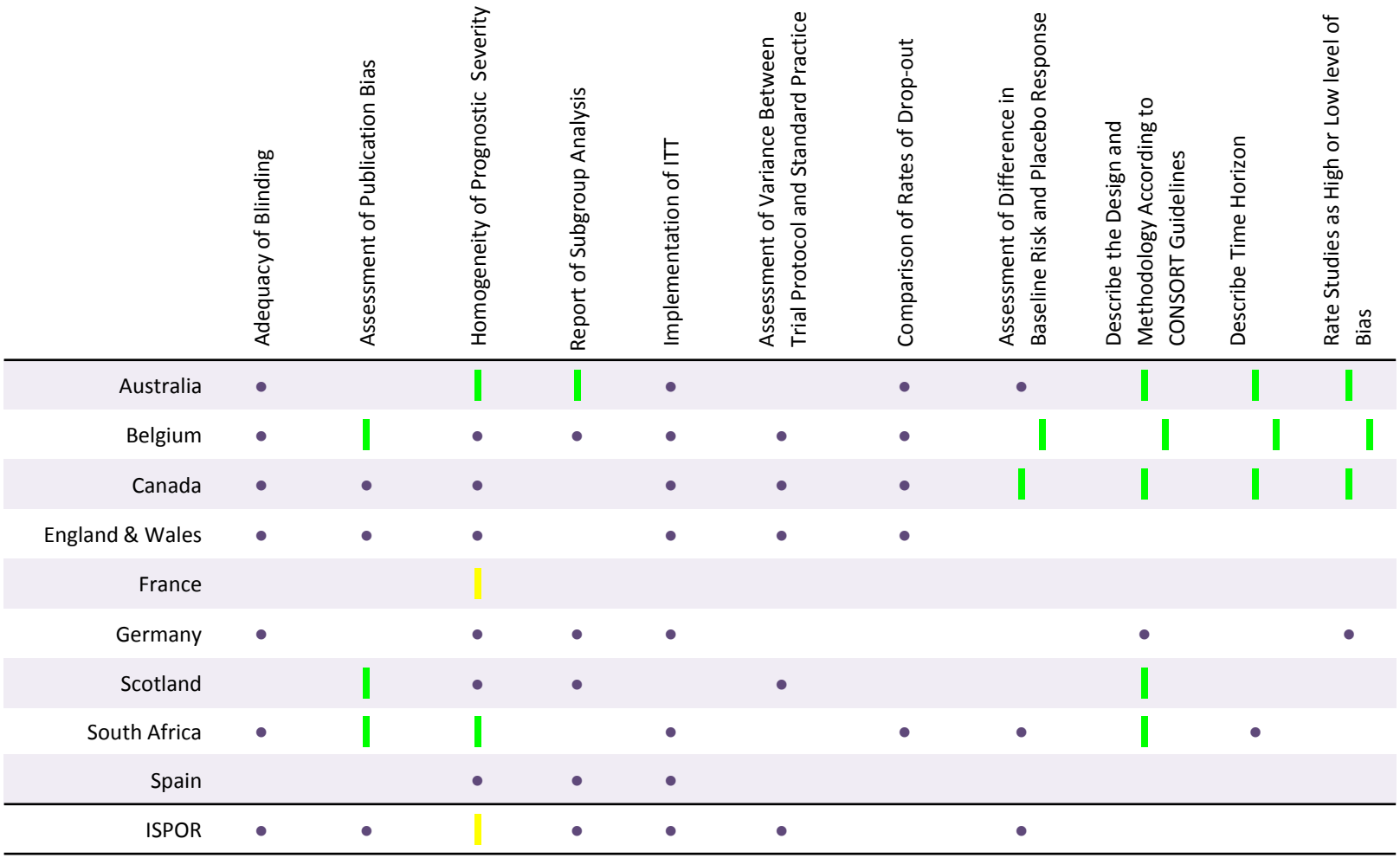
Institute for Quality and Efficiency in Health Care

- Routine use of these methods is not advisable
- In certain situations NMA can be considered
  - Lower certainty of results
- Only accepts adjusted indirect comparisons
  - Bucher, MTC
- Assumption of consistency is critical
- Full description of the model and unclear issues

Recommendations by EUnetHTA on direct and indirect comparisons

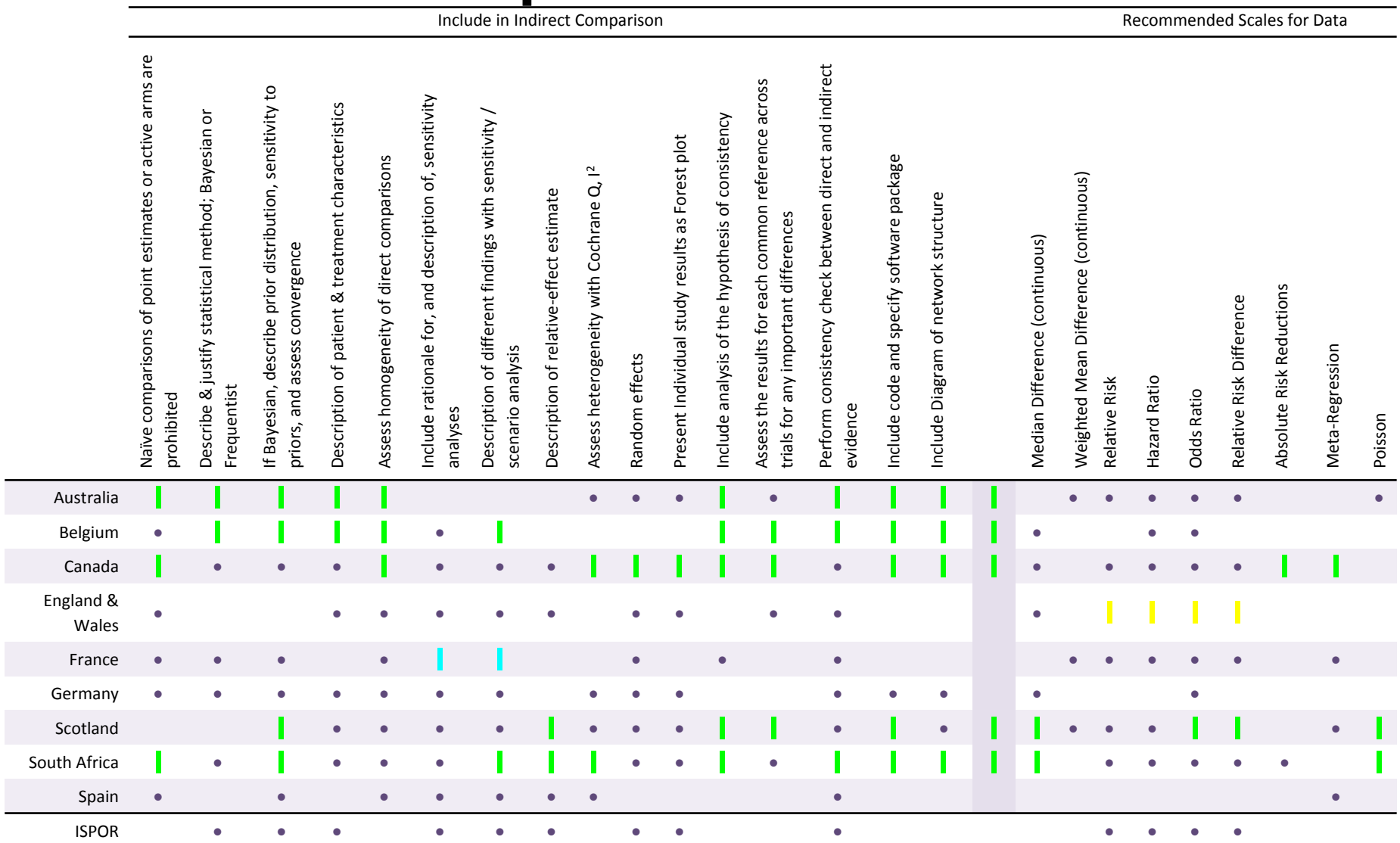
1. Systematic review is a pre-requisite
2. Only combine comparable studies
3. Choice of model (fixed vs random) based on characteristics of studies
4. Investigate potential sources of bias
5. Apply range of sensitivity analyses, e.g. outliers
6. Direct evidence preferred
7. Evaluate direct and indirect evidence separately
8. Use methods that maintain randomisation
9. Choice of method relies on network of evidence
10. Only conduct analyses if data are homogeneous and consistent
11. Explicitly state the assumptions made
12. Justify choice of priors for Bayesian methods
13. Aim for most parsimonious model

Review of national guidelines - bias assessment



A Comparison of National Guidelines for Network Meta-Analysis SUBMITTED FOR PUBLICATION
Andrew Laws MSc(1), Robyn Kendall PGCert BSc(1), Neil Hawkins PhD CStat(2).
(1) Oxford Outcomes (Vancouver), (2) Oxford Outcomes (Oxford)

Review of national guidelines - conduct of indirect comparisons



A Comparison of National Guidelines for Network Meta-Analysis SUBMITTED FOR PUBLICATION
Andrew Laws MSc(1), Robyn Kendall PGCert BSc(1), Neil Hawkins PhD CStat(2).
(1) Oxford Outcomes (Vancouver), (2) Oxford Outcomes (Oxford)

Summary from HTA agency methodology guidelines

- NMAs should only be conducted when RCTs don't exist
- Less weight is given to a NMA compared to RCTs
- Observational data should not be used in a NMA
- Most note that a NMA has relatively low power to detect important differences
- All HTA bodies comment on the underlying assumption that a NMA is only valid if the contributing RCTs are similar

NMA ‘Best Practices’

Neil Hawkins, PhD, CStat
Vice President, Health Economics, Icon PLC
Honorary Professor, (Health Economics and Health Technology Assessment), University of Glasgow

BMJ

RESEARCH

Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews

Fujian Song, reader in research synthesis,[1] Yoon K Loke, senior lecturer in clinical pharmacology,[1] Tanya Walsh, lecturer in dental statistics,[2] Anne-Marie Glenny, lecturer in evidence based oral care,[2] Alison J Eastwood, senior research fellow,[3] Douglas G Altman, professor and director[4]

[1]Faculty of Health, University of East Anglia, Norwich NR4 7TJ
[2]School of Dentistry, University of Manchester, Manchester
[3]Centre for Reviews and Dissemination, University of York, York
[4]Centre for Statistics in Medicine, Oxford

Correspondence to: F Song
fujian.song@uea.ac.uk

Cite this as: *BMJ* 2009;338:b1147
doi:10.1136/bmj.b1147

ABSTRACT

Objective To investigate basic assumptions and other methodological problems in the application of indirect comparison in systematic reviews of competing healthcare interventions.

Design Survey of published systematic reviews.

Inclusion criteria Systematic reviews published between 2000 and 2007 in which an indirect approach had been explicitly used.

Data extraction Identified reviews were assessed for comprehensiveness of the literature search, method for indirect comparison, and whether assumptions about similarity and consistency were explicitly mentioned.

and technology. For many clinical indications clinicians may have to choose among several competing interventions. In this era of evidence based decision making, relative effectiveness and cost effectiveness of different interventions need to be objectively and accurately assessed in clinical studies. It has been accepted generally that well designed and implemented head to head randomised controlled trials provide the most rigorous and valid research evidence on the relative effects of different interventions.[1] Evidence from head to head comparison trials is often limited or unavailable, however, and indirect comparison may therefore be necessary.[2 3]

OPEN ACCESS Freely available online

PLoS one

Indirect Comparisons: A Review of Reporting and Methodological Quality

Sarah Donegan*, Paula Williamson, Carrol Gamble, Catrin Tudur-Smith

Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool, United Kingdom

Abstract

Background: The indirect comparison of two interventions can be valuable in many situations. However, the quality of an indirect comparison will depend on several factors including the chosen methodology and validity of underlying assumptions. Published indirect comparisons are increasingly more common in the medical literature, but as yet, there are no published recommendations of how they should be reported. Our aim is to systematically review the quality of published indirect comparisons to add to existing empirical data suggesting that improvements can be made when reporting and applying indirect comparisons.

Methodology/Findings: Reviews applying statistical methods to indirectly compare the clinical effectiveness of two interventions using randomised controlled trials were eligible. We searched (1966–2008) Database of Abstracts and Reviews of Effects, The Cochrane library, and Medline. Full review publications were assessed for eligibility. Specific criteria to assess quality were developed and applied. Forty-three reviews were included. Adequate methodology was used to calculate the indirect comparison in 41 reviews. Nineteen reviews assessed the similarity assumption using sensitivity analysis, subgroup analysis, or meta-regression. Eleven reviews compared trial-level characteristics. Twenty-four reviews assessed statistical homogeneity. Twelve reviews investigated causes of heterogeneity. Seventeen reviews included direct and indirect evidence for the same comparison; six reviews assessed consistency. One review combined both evidence types. Twenty-five reviews urged caution in interpretation of results, and 24 reviews indicated when results were from indirect evidence by stating this term with the result.

Conclusions: This review shows that the underlying assumptions are not routinely explored or reported when undertaking indirect comparisons. We recommend, therefore, that the quality of indirect comparisons should be improved, in particular, by assessing assumptions and reporting the assessment methods applied. We propose that the quality criteria applied in this article may provide a basis to help review authors carry out indirect comparisons and to aid appropriate interpretation.

Methodological problems in the use of NMA

- Unclear understanding of underlying assumptions
- Incomplete search and inclusion of relevant studies
- Use of flawed or inappropriate methods
- Lack of objective and validated methods to assess or improve trial similarity
- Inadequate comparison and inappropriate combination of direct and indirect evidence

Some NMA ‘Best’ Practice guidelines

- NICE DSU TECHNICAL SUPPORT DOCUMENT 7: EVIDENCE SYNTHESIS OF TREATMENT EFFICACY IN DECISION MAKING: A REVIEWER’S CHECKLIST (7http://bit.ly/HPZS16)
- ISPOR Indirect comparisons and NMA good research practices (http://bit.ly/1aXdapc)
- EUnetHTA: Methodological guideline for REA of pharmaceuticals: Direct and indirect comparison (http://bit.ly/1bJJatz)
- National HTA guidelines: Australia, Belgium, Canada, England & Wales, France, Germany, Scotland, South Africa, Spain

EUnetHTA: Recommendations

1. A systematic literature search is a pre-requisite to conducting a direct or indirect comparison
2. Studies that differ substantially in one or more key characteristics should not be combined
3. The choice between a fixed and random effects model should be based on the characteristics of the studies being analysed
4. Potential sources of bias should be investigated
5. Attention should be given to determining the presence of outliers or influential observations
6. Where sufficient good quality head-to-head studies are available, direct comparisons are preferred as the level of evidence is high
7. If both direct and indirect evidence are available, they can be evaluated separately
8. Only adjusted methods of indirect comparison that maintain randomisation should be used
9. The choice of indirect comparison method relies on the network of available evidence. Preference should be given to the most transparent method available (i.e. one should favour Bucher's method of adjusted indirect comparison over MTC if the data permit its usage and the appropriate assumptions are satisfied)
10. An indirect comparison should only be carried out if underlying data from comparable studies are homogeneous and consistent, otherwise the results will not be reliable
11. The assumptions made for indirect comparisons must be explicitly stated. Particular attention should be given to the homogeneity, similarity and consistency assumptions. A general assumption of indirect comparisons is that the relative effectiveness of a treatment is the same across all studies included in a meta-analysis
12. When Bayesian methods are applied, the choice of the prior distributions should be justified and documented
13. The complexity of a model is not a measure of its accuracy or utility and preference is for the most parsimonious model whose assumptions can be justified

Enron's accounts

	Dollar amounts in millions				
	2000	1999	1998	1997	1996
Summary data for unconsolidated affiliates					
Revenues	15,903	11,568	8,508	11,183	11,676
Net income	586	1,857	142	336	464
Current assets	5,884	3,168	2,309	3,611	2,587
Total assets	34,155	26,983	22,125	8,851	8,064
Current liabilities	4,739	4,401	3,501	1,089	902
Total liabilities	20,604	15,289	13,138	13,551	11,553
Owners' equity	13,551	11,694	8,987	1,861	2,381
Equity in earnings of affiliates	87	309	97	216	215
Pro forma *ratio analysis – as if consolidated*					
Current assets/current liabilities	1.09	0.93	0.86	1.51	1.42
Total liabilities/stockholders' equity	2.98	1.84	2.21	4.19	3.93
Net profit margin	0.84%	1.73%	1.77%	1.64%	1.97%
Asset turnover	1.17	0.86	0.77	0.97	1.03
Return on assets	0.98%	1.48%	1.37%	1.60%	2.04%
Assets/stockholders' equity	3.98	2.84	3.21	4.32	3.96
Return on stockholders' equity	3.91%	4.20%	4.38%	6.89%	8.08%

Source: Enron Corp. Annual Reports

What is the link to network meta-analysis?

…like accounts, interpretation of NMAs requires judgement (plus we should worry about off-balance sheet items)

Accounting standards

- U.S. Generally Accepted Accounting Principles(US GAAP): rule-based
- International Financial Reporting Standards (IFRS): principle-based

Network meta-analysis

- What are the principles?

Perhaps?

- Should systematically identify and use relevant evidence
- Help consumers to evaluate the usefulness of the analysis (based on the consistency constraint)
- Report transparently (in sufficient detail to allow replication and modification)

Identify and use relevant evidence

- Define PICO criteria (Population, Intervention, Comparators, Outcomes) correspond to decision problem
- May need to extend search beyond treatments in PICO criteria
- May include unlicensed treatments
- Need to conduct a credible and repeatable search (PRISMA guidelines)

May need to extend search beyond treatments in PICO criteria

How Far Do You Go? Efficient Searching for Indirect Evidence

Neil Hawkins, PhD, MSc, David A. Scott, MA, Beth Woods, BA

Background. *Indirect evidence is particularly valuable in health care decision making when direct trial evidence comparing relevant treatments is absent or limited. Current approaches using a predetermined set of comparators in the search query may fail to identify all relevant indirect evidence.* ***Purpose.*** *To present a framework for the efficient design of search strategies for identifying clinical trials providing indirect evidence for a treatment comparison.* ***Findings.*** *The authors present 2 search strategies that differ from traditional search strategies in using a series of iterative searches to identify the set of relevant comparators. In both, the comparators included in each search are determined by the results of previous searches. For a given number of searches, the strategies presented will find all indirect comparisons that include a certain number of comparators linking the treatments of interest. Methods of estimating the value of indirect evidence via a given number of comparators linking the treatments of interest are presented, thus allowing the burden of additional searching to be traded off against the likely impact of finding more distant comparisons. A practical illustration of the search strategies in the context of informing a network meta-analysis of second-line treatments for non-small cell lung cancer is presented.* ***Conclusions.*** *The iterative strategies presented offer a means of identifying such evidence and allow the researcher to determine the optimal scope of the search by estimating the value of additional indirect evidence.* ***Key words:*** *evidence synthesis; indirect evidence; search strategies; oncology.* ***(Med Decis Making 2009;29:273–281)***

Help consumers to evaluate the usefulness of the analysis

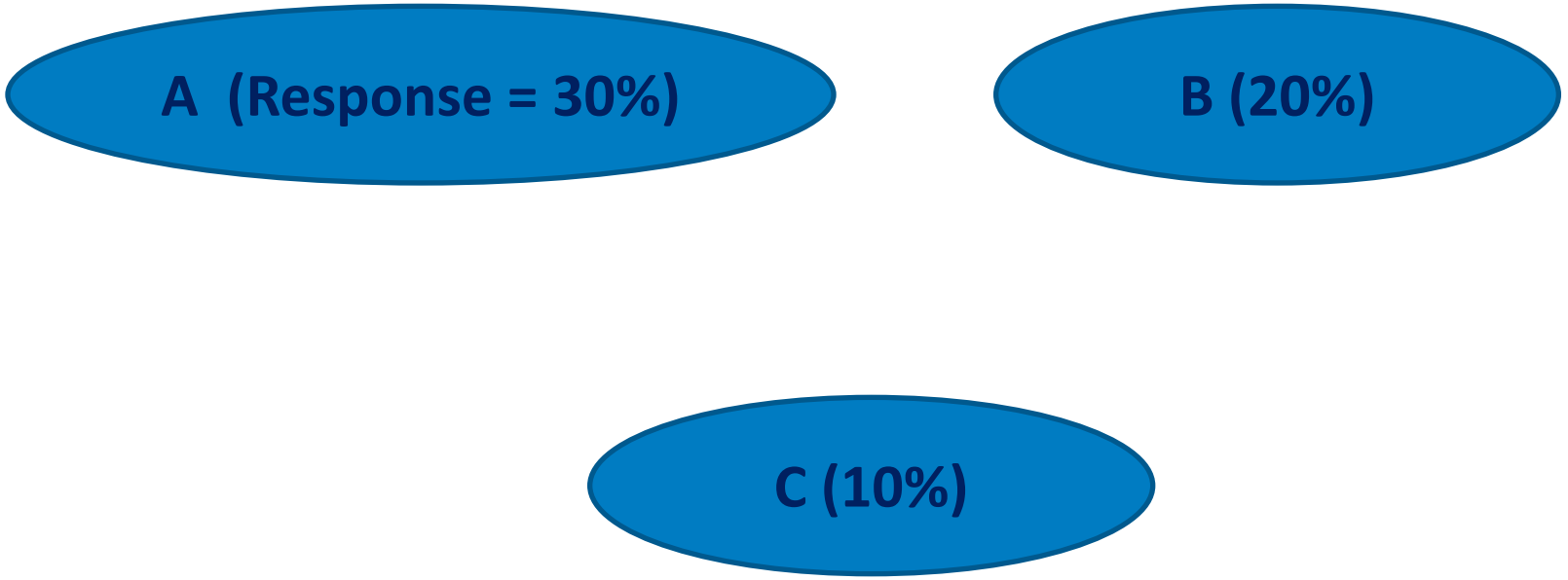
$$\partial_{AB} = \partial_{AC} - \partial_{BC}$$

Referred to as:

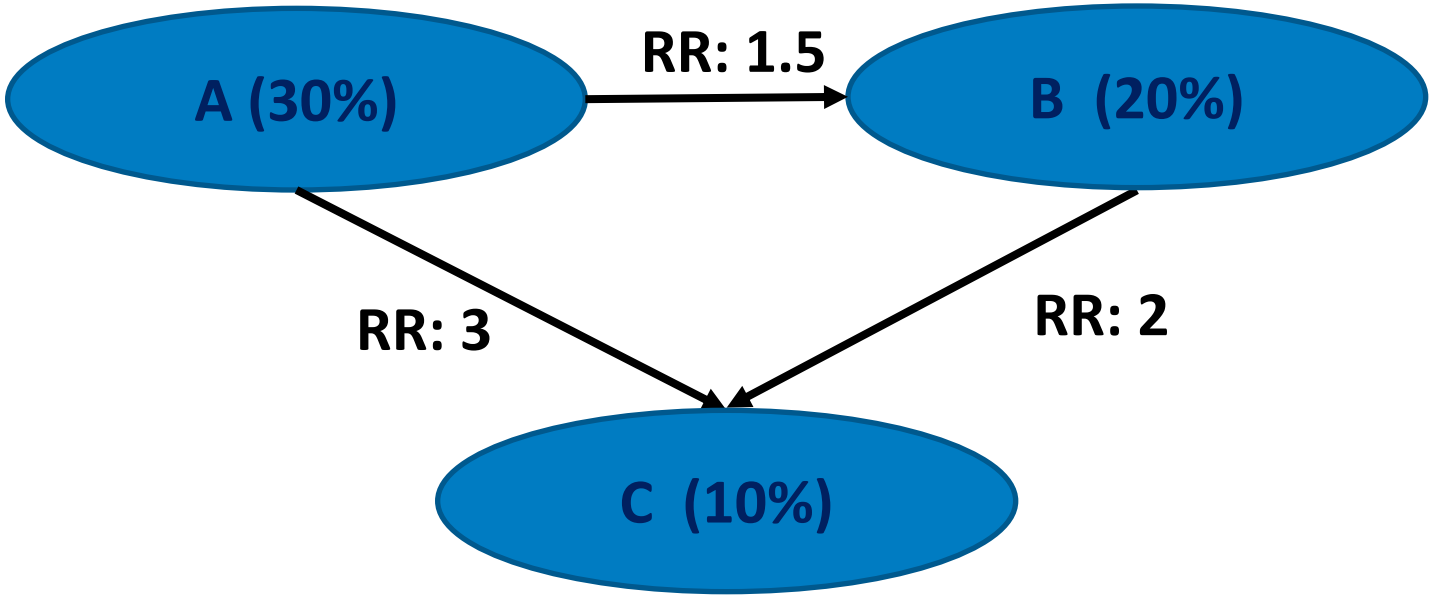
- Consistency
  - Indirect and direct estimates are consistent
- Exchangeability
  - If treatments were exchanged between trials estimates would be the same (allowing for random variation)
- Similarity
  - The trials are similar and comparable
- Transitivity

$$\partial_{AB} = \partial_{AC} - \partial_{BC} \qquad \partial_{AC} = \partial_{AB} - \partial_{CB}$$

Consider a single trial



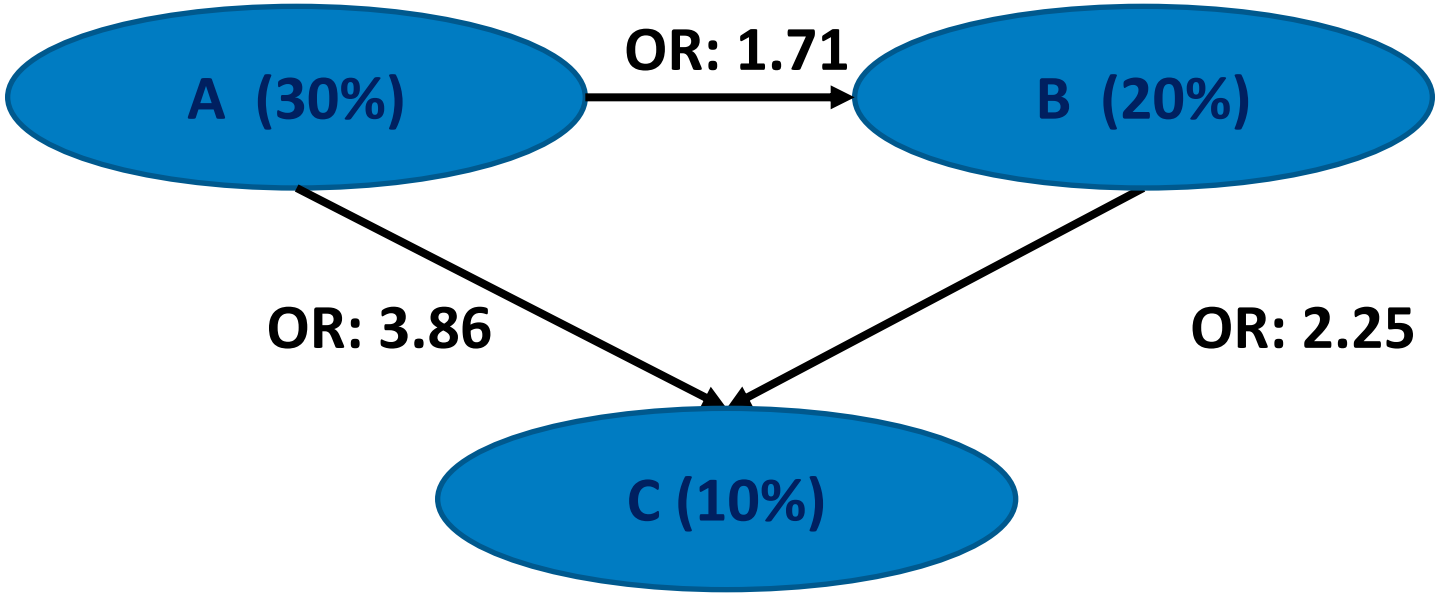
By definition consistent on the relative risk scale...



$$RR_{AvsB} = {RR_{AvsC}}/{RR_{BvsC}}$$

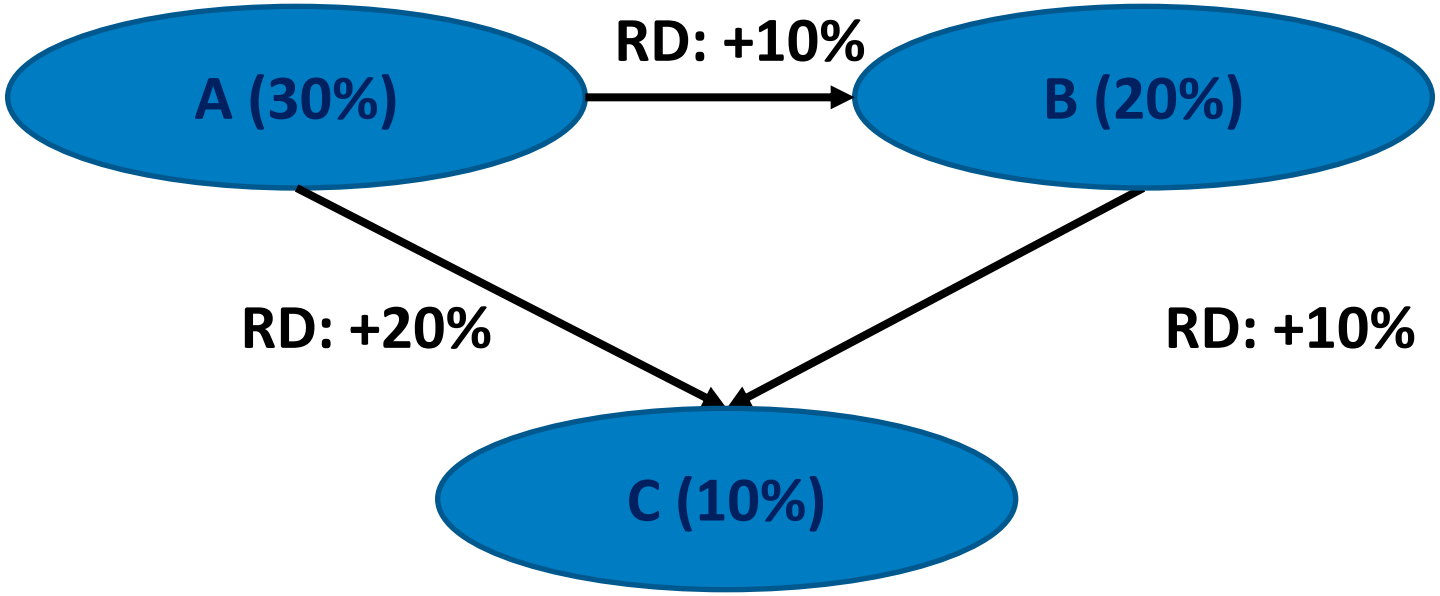
$$RR_{AvsB} = {3}/{2} = 1.5$$

Also on the odds ratio scale...



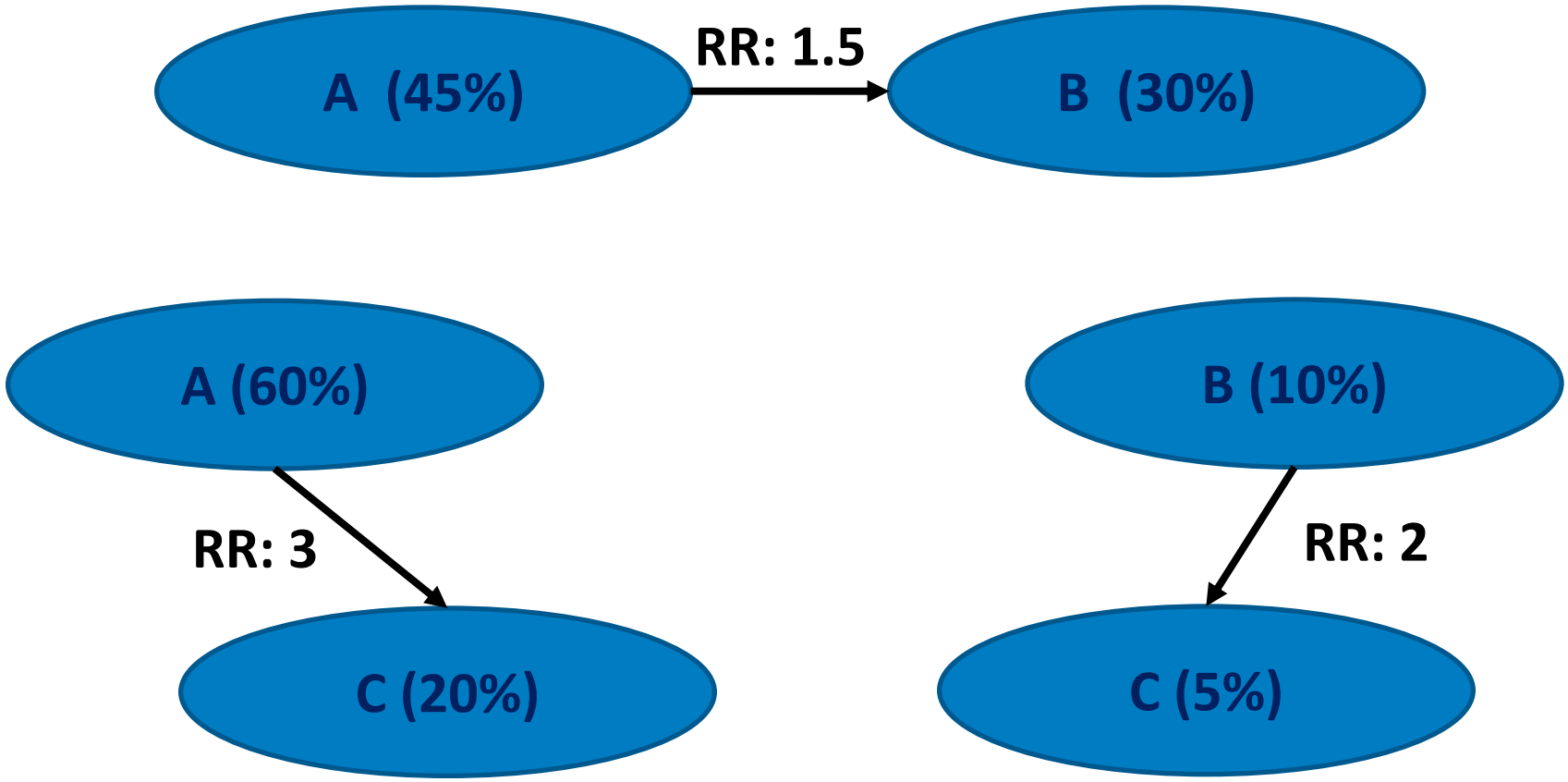
$$OR_{AvsB} = {OR_{AvsC}}/{OR_{BvsC}} \qquad OR_{AvsB} = {3.86}/{2.25} = 1.71$$

And on the risk difference (RD) scale...



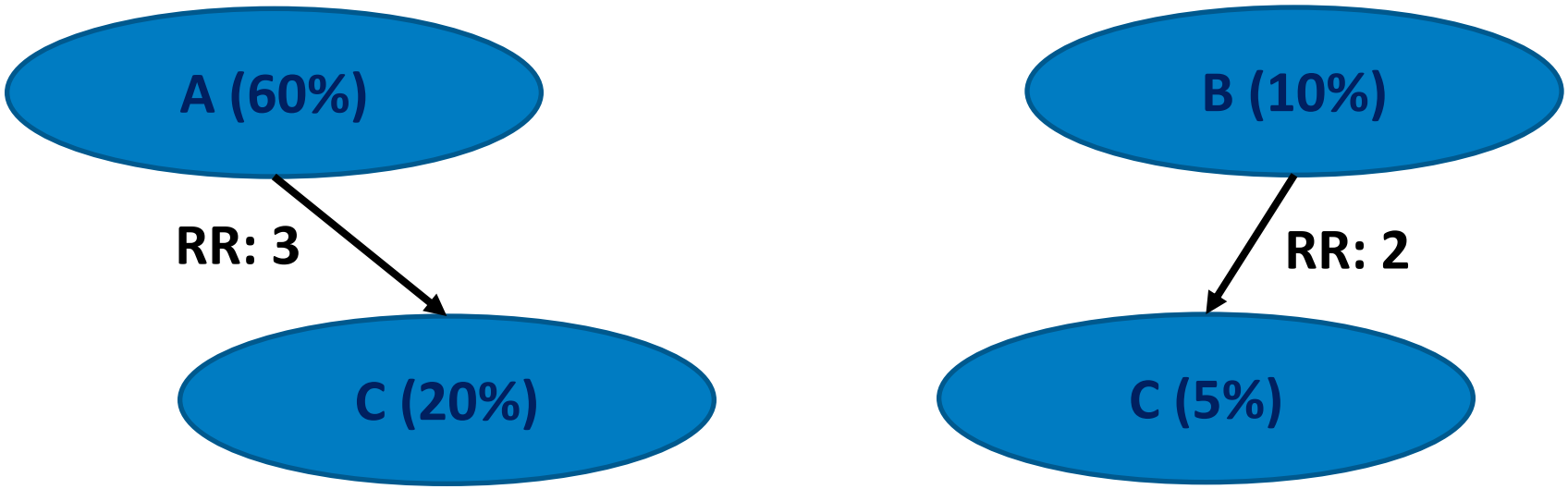
$$RD_{AvsB} = RD_{AvsC} - RD_{BvsC} \qquad RD_{AvsB} = 20\% - 10\% = +10\%$$

Consider multiple trials

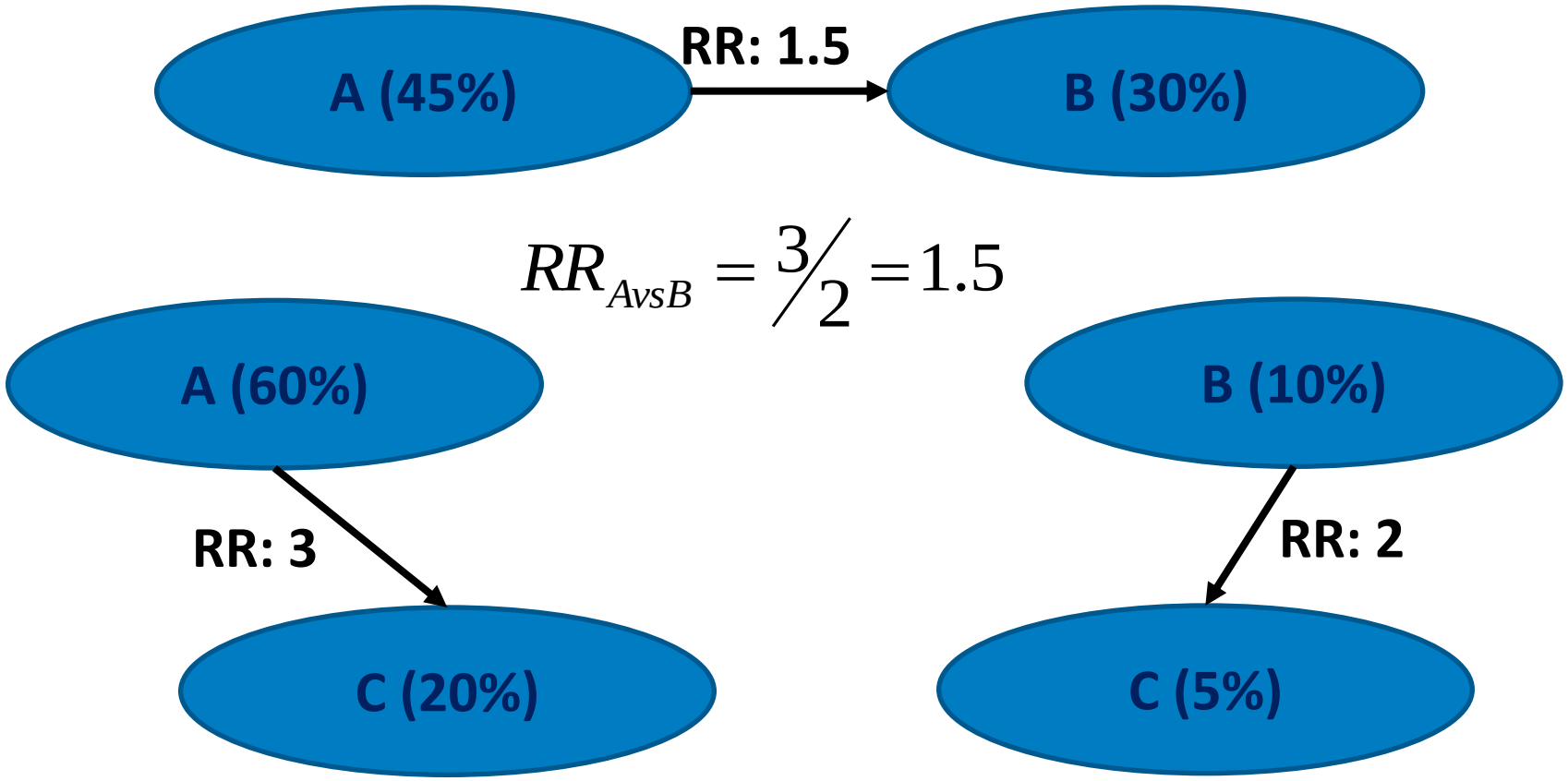


The indirect estimate of RR A vs. B…

$$RR_{AvsB} = {RR_{AvsC}}/{RR_{BvsC}} = {3}/{2} = 1.5$$

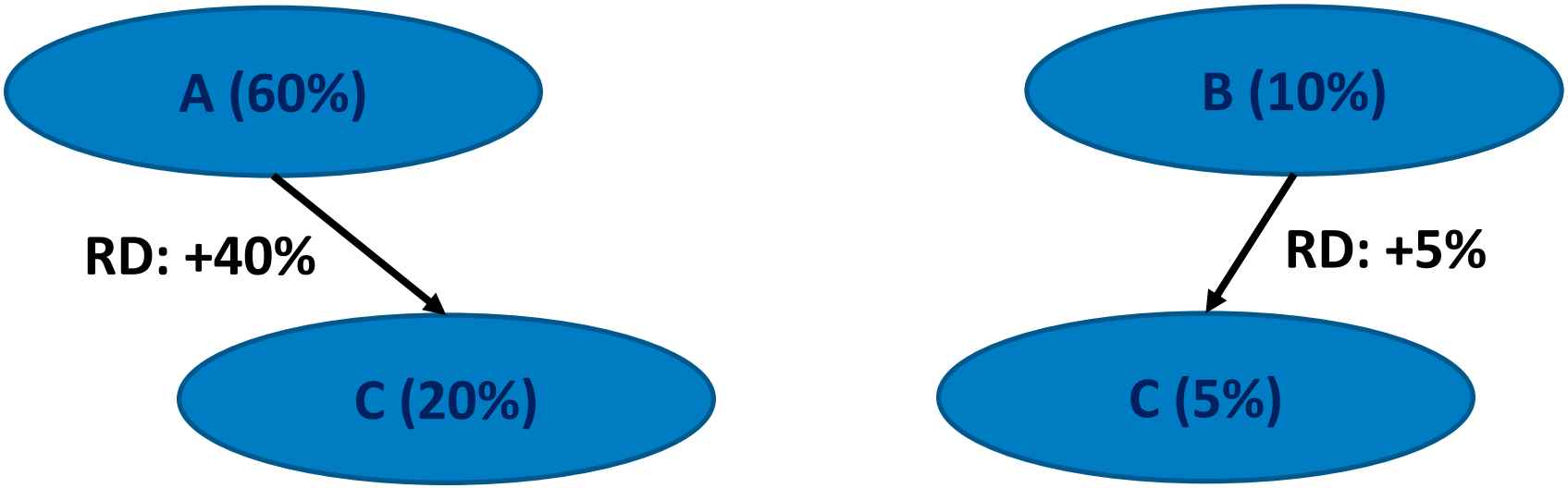


Is consistent with the direct estimate…

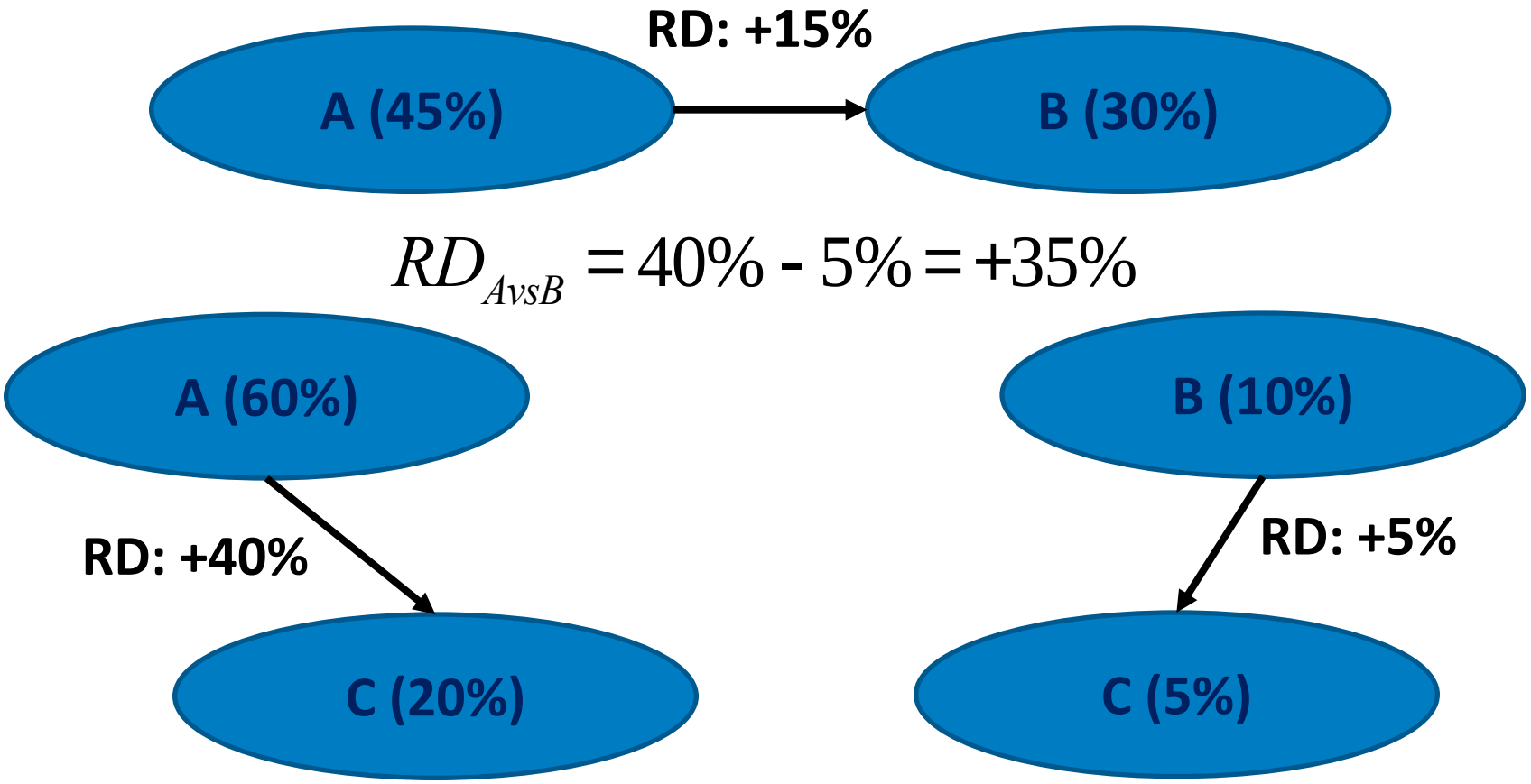


The indirect estimate of RD A vs. B

$$RD_{AvsB} = RD_{AvsC} - RD_{BvsC} = 40\% - 5\% = 35\%$$



Is inconsistent with the direct estimate…



Consistency

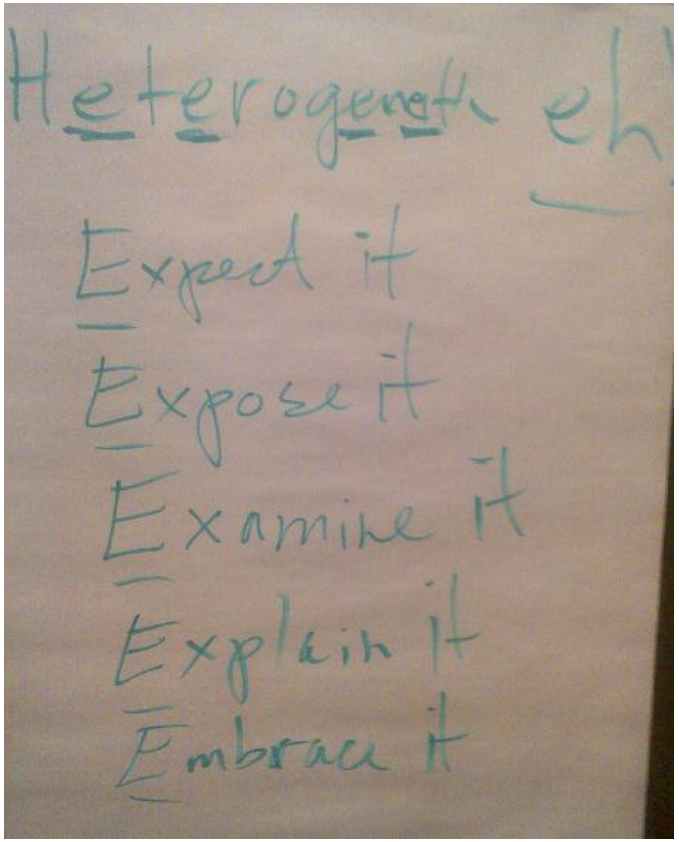
- Is only an assumption, not a natural law
- Randomisation within trials supports internal
- Comparisons across trials are observational
- May depend on choice of scale
- Is a ‘model’

"essentially, all models are wrong, but some are useful" George Box

Need to consider heterogeneity

Heterogeneity eh!

- Expect it
- Expose it
- Examine it
- Explain it
- Embrace it



George Wells

Expect it!

- Differences in patients
- Differences in study designs
- Differences in treatments

Examples?

- Comparison of average treatment effect estimates
  - Biased by predictive factors
  - Not biased by ‘purely’ prognostic factors (on the scale used for analysis)

Expose it!

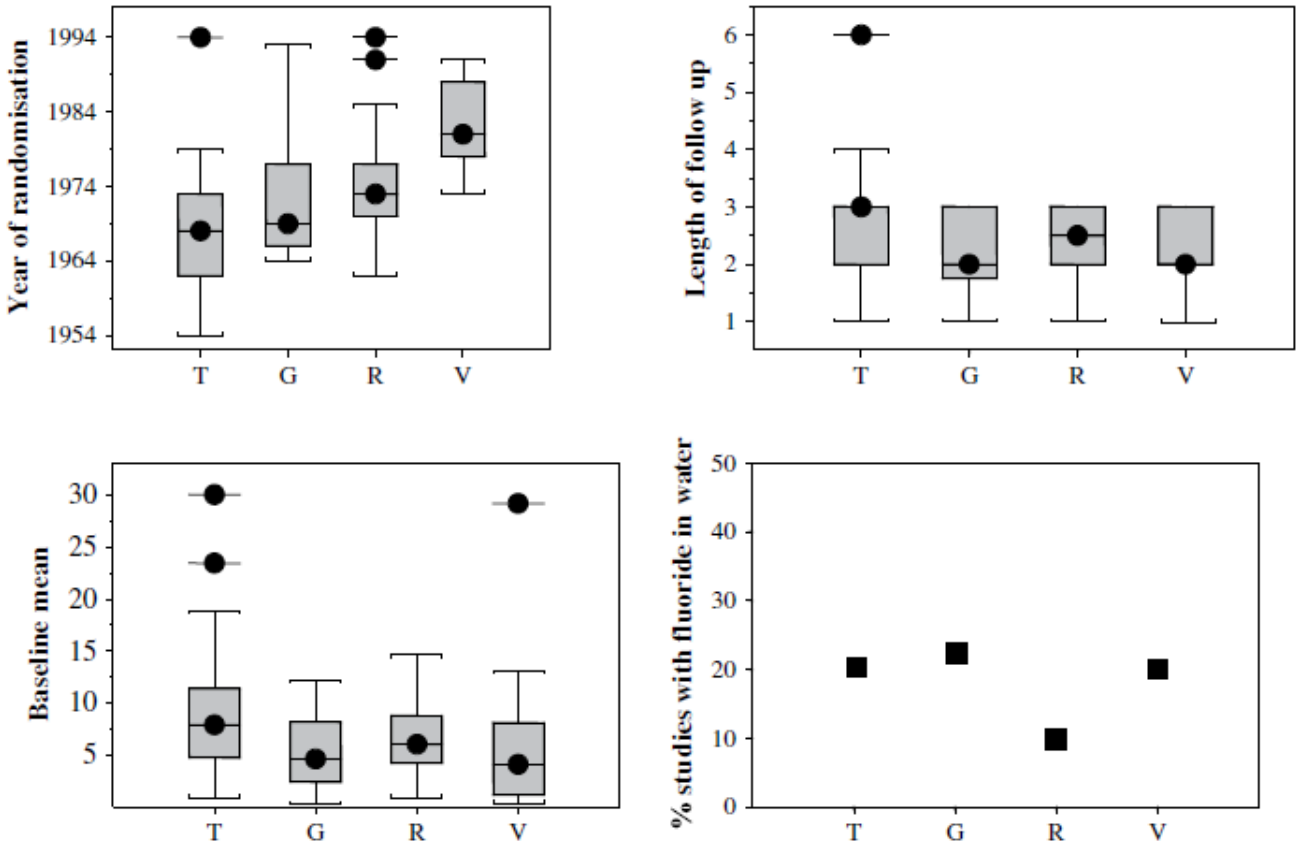


Fig. 4. Distribution of the possible confounding factors: box plots for year of randomization, length of follow-up, and baseline mean caries; percentage of the studies carried out in populations with fluoridation in the water.

Examine it!

Table III. Smoking cessation: posterior means and standard deviations (sd) of the log-odds ratios calculated using all the evidence (MTC) and when direct and indirect evidence on each node is split and posterior mean and sd of the inconsistency estimate (calculated as direct–indirect at each iteration) with Bayesian *p*-value, *P*, measuring agreement between direct and indirect evidence for each split node.

Treatments		MTC		Direct		Indirect		Inconsistency estimate		
X	*Y*	Mean	sd	Mean	sd	Mean	sd	Mean	sd	*P*
A	B	0.493	0.406	0.342	0.55	0.706	0.635	−0.365	0.840	0.65
A	C	0.844	0.240	0.845	0.254	0.673	0.679	0.171	0.716	0.79
A	D	1.106	0.442	1.360	0.829	1.108	0.539	0.253	0.983	0.81
B	C	0.352	0.416	−0.052	0.702	0.519	0.503	−0.571	0.853	0.49
B	D	0.613	0.488	0.676	0.698	0.511	0.684	0.165	0.966	0.85
C	D	0.261	0.419	−0.085	0.479	1.708	0.893	−1.793	1.009	0.07

Statist. Med. ***2010***, 29 932–944

Assessing the effectiveness of primary angioplasty compared with thrombolysis and its relationship to time delay: a Bayesian evidence synthesis

Christian Asseburg, Yolanda Bravo Vergel, Stephen Palmer, Elisabeth Fenwick, Mark de Belder, Keith R Abrams, Mark Sculpher

Heart 2007;**93**:1244–1250. doi: 10.1136/hrt.2006.093336

A supplementary technical report is available on the *Heart* website—http://heart.bmj.com/supplemental

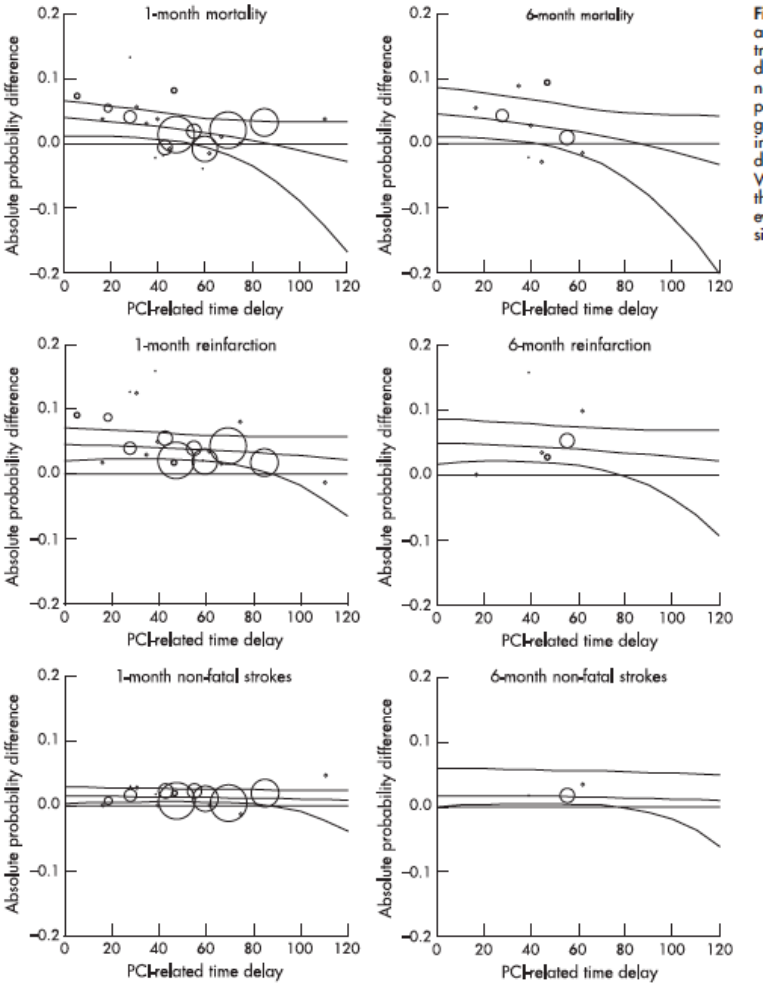
See end of article for authors' affiliations

Correspondence to: Dr M Sculpher, Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK; mjs23@york.ac.uk

Accepted 19 December 2006
Published Online First 3 February 2007

Background: Meta-analyses of trials have shown greater benefits from angioplasty than thrombolysis after an acute myocardial infarction, but the time delay in initiating angioplasty needs to be considered.
Objective: To extend earlier meta-analyses by considering 1- and 6-month outcome data for both forms of reperfusion. To use Bayesian statistical methods to quantify the uncertainty associated with the estimated relationships.
Methods: A systematic review and meta-analysis published in 2003 was updated. Data on key clinical outcomes and the difference between time-to-balloon and time-to-needle were independently extracted by two researchers. Bayesian statistical methods were used to synthesise evidence despite differences between reported follow-up times and outcomes. Outcomes are presented as absolute probabilities of specific events and odds ratios (ORs; with 95% credible intervals (CrI)) as a function of the additional time delay associated with angioplasty.
Results: 22 studies were included in the meta-analysis, with 3760 and 3758 patients randomised to primary angioplasty and thrombolysis, respectively. The mean (SE) angioplasty-related time delay (over and above time to thrombolysis) was 54.3 (2.2) minutes. For this delay, mean event probabilities were lower for primary angioplasty for all outcomes. Mortality within 1 month was 4.5% after angioplasty and 6.4% after thrombolysis (OR = 0.68 (95% CrI 0.46 to 1.01)). For non-fatal reinfarction, OR = 0.32 (95% CrI 0.20 to 0.51); for non-fatal stroke OR = 0.24 (95% CrI 0.11 to 0.50). For all outcomes, the benefit of angioplasty decreased with longer delay from initiation.
Conclusions: The benefit of primary angioplasty, over thrombolysis, depends on the former's additional time delay. For delays of 30–90 minutes, angioplasty is superior for 1-month fatal and non-fatal outcomes. For delays of around 90 minutes thrombolysis may be the preferred option as assessed by 6-month mortality; there is considerable uncertainty for longer time delays.

Explain it!


1-month mortality
6-month mortality
1-month reinfarction
6-month reinfarction
1-month non-fatal strokes
6-month non-fatal strokes
Absolute probability difference
PCI-related time delay
0.2
0.1
0.0
−0.1
−0.2
0
20
40
60
80
100
120

INTERVENTIONAL CARDIOLOGY

Is primary angioplasty cost effective in the UK? Results of a comprehensive decision analysis

Yolanda Bravo Vergel, Stephen Palmer, Christian Asseburg, Elisabeth Fenwick, Mark de Belder, Keith Abrams, Mark Sculpher

UNLOCKED

This paper is freely available online under the BMJ Journals unlocked scheme, see http://heart.bmj.com/info/unlocked.dtl

Heart 2007;**93**:1238–1243. doi: 10.1136/hrt.2006.111401

See end of article for authors' affiliations

Correspondence to: Dr M Sculpher, Centre for Health Economics, University of York, Heslington, YO10 5DD York, UK; mjs23@york.ac.uk

Accepted 23 May 2007

Objective: To assess the cost effectiveness of primary angioplasty, compared with medical management with thrombolytic drugs, to achieve reperfusion after acute myocardial infarction (AMI) from the perspective of the UK NHS.

Design: Bayesian evidence synthesis and decision analytic model.

Methods: A systematic review was conducted and Bayesian statistical methods used to synthesise evidence from 22 randomised control trials. Resource utilisation was based on UK registry data, published literature and national databases, with unit costs taken from routine NHS sources and published literature.

Main outcome measure: Costs from a health service perspective and outcomes measured as quality-adjusted life years (QALYs).

Results: For the base case, the incremental cost-effectiveness ratio of primary angioplasty was £9241 for each additional QALY, with a probability of being cost effective of 0.90 for a cost-effectiveness threshold of £20 000. Results were sensitive to variations in the additional time required to initiate treatment with primary angioplasty.

Conclusions: Primary angioplasty is cost effective for the treatment of AMI on the basis of threshold cost-effectiveness values used in the NHS and subject to a delay of up to about 80 minutes. These findings are mainly explained by the superior mortality benefit and the prevention of non-fatal outcomes associated with primary angioplasty for delays of up to this length.

Embrace it!

Table 3 Cost-effectiveness results

Time delay	Strategy	Mean costs (£)	Mean QALYs	ICER (£)	Probability of being cost effective for threshold of: £10 000	£20 000	£30 000
Base-case analysis							
Average delay (54 min)	Primary angioplasty	12 760	7.12	9 241	0.55	0.90	0.95
	Thrombolysis	10 080	6.83	–	0.45	0.10	0.05
Time delays of:							
30 Minutes	Primary angioplasty	12 820	7.23	6 850	0.82	0.98	0.99
	Thrombolysis	10 080	6.83	–	0.18	0.02	0.01
60 Minutes	Primary angioplasty	12 750	7.09	10 269	0.43	0.83	0.91
	Thrombolysis	10 080	6.83	–	0.57	0.17	0.09
90 Minutes	Primary angioplasty	12 670	6.87	64 750	0.13	0.36	0.45
	Thrombolysis	10 080	6.83	–	0.87	0.64	0.55
Differential length of hospital stay*							
Average delay (54 min)	Primary angioplasty	12 030	7.12	5 448	0.82	0.95	0.97
	Thrombolysis	10 450	6.83	–	0.18	0.05	0.03
30 Minutes	Primary angioplasty	12 085	7.23	4 087	0.95	0.99	0.99
	Thrombolysis	10 450	6.83	–	0.05	0.01	0.01
60 Minutes	Primary angioplasty	12 020	7.09	6 038	0.75	0.91	0.99
	Thrombolysis	10 450	6.83	–	0.25	0.09	0.01
90 Minutes	Primary angioplasty	11 940	6.87	37 250	0.32	0.47	0.52
	Thrombolysis	10 450	6.83	–	0.68	0.53	0.48

ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years.
*Using an estimate of 5.8 days (SE 1.6) for primary angioplasty and 12.1 days (SE 2.9) for thrombolysis (Morgan K, personal communication, 2005).

‘Studies that differ substantially in one or more key characteristics (e.g. participants, interventions, outcomes measured) should not be combined’

Discuss?

Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions☆

F. Song[a,b,*], I. Harvey[a], R. Lilford[c]

[a]*School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK*
[b]*School of Allied Health Professions, University of East Anglia, Norwich, NR4 7TJ, UK*
[c]*Department of Public Health and Epidemiology, University of Birmingham, Birmingham, B15 2TT, UK*

Accepted 5 June 2007

Abstract

Objective: To investigate discrepancies between direct comparison and adjusted indirect comparison in meta-analyses of new versus conventional pharmaceutical interventions.

Study Design and Setting: Results of direct comparison were compared with results of adjusted indirect comparison in three meta-analyses of new versus conventional drugs. The three case studies are (1) bupropion versus nicotine replacement therapy for smoking cessation, (2) risperidone versus haloperidol for schizophrenia, and (3) fluoxetine versus imipramine for depressive disorders.

Results: In all the three cases, effects of new drugs estimated by head-to-head trials tend to be greater than that by adjusted indirect comparisons. The observed discrepancies could not be satisfactorily explained by the play of chance or by bias and heterogeneity in adjusted indirect comparison. This observation, along with analysis of possible systematic bias in the direct comparisons, suggested that the indirect method might have produced less biased results. Simulations found that adjusted indirect comparison may counterbalance bias under certain circumstances.

Research Article

Statistics in Medicine

Received 15 December 2009, Accepted 26 May 2010 Published online 4 August 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4001

Evaluating novel agent effects in multiple-treatments meta-regression

Georgia Salanti,[a*†] Sofia Dias,[b] Nicky J. Welton,[b] A. E. Ades,[b] Vassilis Golfinopoulos,[c] Maria Kyrgiou,[d] Davide Mauri[a,e] and John P. A. Ioannidis[a,f,g]

Multiple-treatments meta-analyses are increasingly used to evaluate the relative effectiveness of several competing regimens. In some fields which evolve with the continuous introduction of new agents over time, it is possible that in trials comparing older with newer regimens the effectiveness of the latter is exaggerated. Optimism bias, conflicts of interest and other forces may be responsible for this exaggeration, but its magnitude and impact, if any, needs to be formally assessed in each case. Whereas such novelty bias is not identifiable in a pair-wise meta-analysis, it is possible to explore it in a network of trials involving several treatments. To evaluate the hypothesis of novel agent effects and adjust for them, we developed a multiple-treatments meta-regression model fitted within a Bayesian framework. When there are several multiple-treatments meta-analyses for diverse conditions within the same field/specialty with similar agents involved, one may consider either different novel agent effects in each meta-analysis or may consider the effects to be exchangeable across the different conditions and outcomes. As an application, we evaluate the impact of modelling and adjusting for novel agent effects for chemotherapy and other non-hormonal systemic treatments for three malignancies. We present the results and the impact of different model assumptions to the relative ranking of the various regimens in each network. We established that multiple-treatments meta-regression is a good method for examining whether novel agent effects are present and estimation of their magnitude in the three worked examples suggests an exaggeration of the hazard ratio by 6 per cent (2–11 per cent). **Copyright © 2010 John Wiley & Sons, Ltd.**

Keywords: network meta-analysis; novelty bias; optimism bias; mixed-treatment comparison

ELSEVIER

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jval

Consistency between Direct and Indirect Trial Evidence: Is Direct Evidence Always More Reliable?

Jason Madan, MA, MSc, PhD, Matt D. Stevenson, BSc, PhD, Katy L. Cooper, BSc, PhD, A. E. Ades, PhD, Sophie Whyte, BSc, MMath, PhD, Ron Akehurst, BSc (Econ), Hon MFPH*

University of Bristol, Bristol, UK

A B S T R A C T

Objectives: To present a case study involving the reduction in incidence of febrile neutropenia (FN) after chemotherapy with granulocyte colony–stimulating factors (G-CSFs), illustrating difficulties that may arise when following the common preference for direct evidence over indirect evidence. **Methods:** Evidence of the efficacy of treatments was identified from two previous systematic reviews. We used Bayesian evidence synthesis to estimate relative treatment effects based on direct evidence, indirect evidence, and both pooled together. We checked for inconsistency between direct and indirect evidence and explored the role of one specific trial using cross-validation. A subsequent review identified further studies not available at the time of the original analysis. We repeated the analyses on the enlarged evidence base. **Results:** We found substantial inconsistency in the original evidence base. The median odds ratio of FN for primary pegfilgrastim versus no primary G-CSF was 0.06 (95% credible interval: 0.02–0.19) based on direct evidence, but 0.27 (95% credible interval: 0.13–0.53) based on indirect evidence (P value for consistency hypothesis 0.027). The additional trials were consistent with the earlier indirect, rather than the direct, evidence, and there was no inconsistency between direct and indirect estimates in the updated evidence. The earlier inconsistency was due to one trial comparing primary pegfilgrastim with no primary G-CSF. Predictive cross-validation showed that this study was inconsistent with the evidence as a whole and with other trials making this comparison. **Conclusions:** Both the Cochrane Handbook and the NICE Methods Guide express a preference for direct evidence. A more robust strategy, which is in line with the accepted principles of evidence synthesis, would be to combine all relevant and appropriate information, whether direct or indirect.

Keywords: Bayesian methods, febrile neutropenia, granulocyte colony–stimulating factors, methodology, mixed treatment comparison

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Special Issue Paper

Research Synthesis Methods

Received 10 August 2011, Revised 1 February 2012, Accepted 17 February 2012 Published online 1 June 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.57

Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions‡

Anna Chaimani and Georgia Salanti*†

Suggested methods for exploring the presence of small-study effects in a meta-analysis and the possibility of publication bias are associated with important limitations. When a meta-analysis comprises only a few studies, funnel plots are difficult to interpret, and regression-based approaches to test and account for small-study effects have low power. Assuming that the cause of funnel plot asymmetry is likely to affect an entire research field rather than only a particular comparison of interventions, we suggest that network meta-regression is employed to account for small-study effects in a set of related meta-analyses. We present several possible models for the direction and distribution of small-study effects and we describe the methods by re-analysing two published networks. **Copyright © 2012 John Wiley & Sons, Ltd.**

Keywords: funnel plot; publication bias; sponsorship bias; optimism bias; selective reporting bias

‘Where sufficient good quality head-to-head studies are available, direct comparisons are preferred as the level of evidence is high’

Discuss?

Transparent Reporting

Original Research

Journal of Health Services Research & Policy
18(4) 224–232
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1355819613498379
jhsrp.rsmjournals.com
SAGE

Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons

Sze Huey Tan[1], Sylwia Bujkiewicz[2], Alexander Sutton[3], Pascale Dequen[4] and Nicola Cooper[5]

Abstract

Objectives: To establish current guidance and practice in UK on presentation of indirect comparison and mixed treatment comparison analyses; to provide recommendations to improve indirect comparison/mixed treatment comparison reporting and to identify research priorities for improved presentation.

Methods: Existing institutional guidance for conducting indirect comparison/mixed treatment comparison alongside current practice in health technology assessment was reviewed. Reports published in UK by the Health Technology Assessment programme since 1997, which utilized indirect comparison/mixed treatment comparison methods, were reviewed with respect to the presentation of study data, statistical models and results. Recommendations for presentation were developed.

Results: Guidance exists that provide the details necessary to conduct a successful indirect comparison/mixed treatment comparison analysis but recommendations on presentation are limited. Of 205 health technology assessment reports that contained evidence synthesis for effectiveness, 19 used indirect comparison/mixed treatment comparison methods. These reports utilized numerous presentational formats from which the following key components were identified: network table/diagram for presenting data; model description to allow reproducibility; and tables, forest plots, matrix tables and summary forest plots for presenting a range of results. Recommendations were developed to ensure that reporting is explicit, transparent and reproducible. Approaches most understandable by non-technical decision makers, and areas where future research is required, are outlined.

Conclusions: There is no standard presentational style used in UK for reporting indirect comparison/mixed treatment comparison, and the use of graphical tools is limited. Standardization of reporting and innovation in graphical representation of indirect comparison/mixed treatment comparison results is required.

Keywords

indirect treatment comparisons, mixed treatment comparisons, reporting

Recommendations

- Clearly identify studies
- Report treatments included in each study
- Report data / effect sizes for each study
- Describe statistical model
- Supply code and data
- Report treatment effects compared to reference treatment
- Report pairwise comparison of all treatments
- Report probability best / ranking of treatments

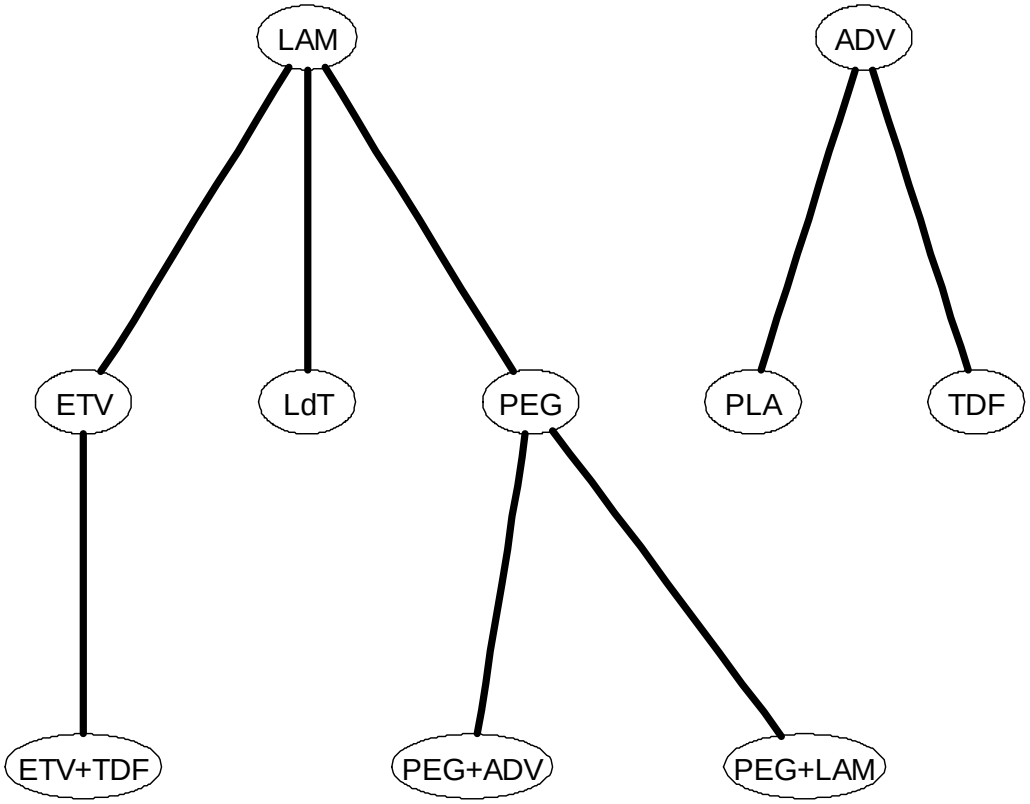
GASTROENTEROLOGY 2010;139:1218–1229

Tenofovir and Entecavir Are the Most Effective Antiviral Agents for Chronic Hepatitis B: A Systematic Review and Bayesian Meta-analyses

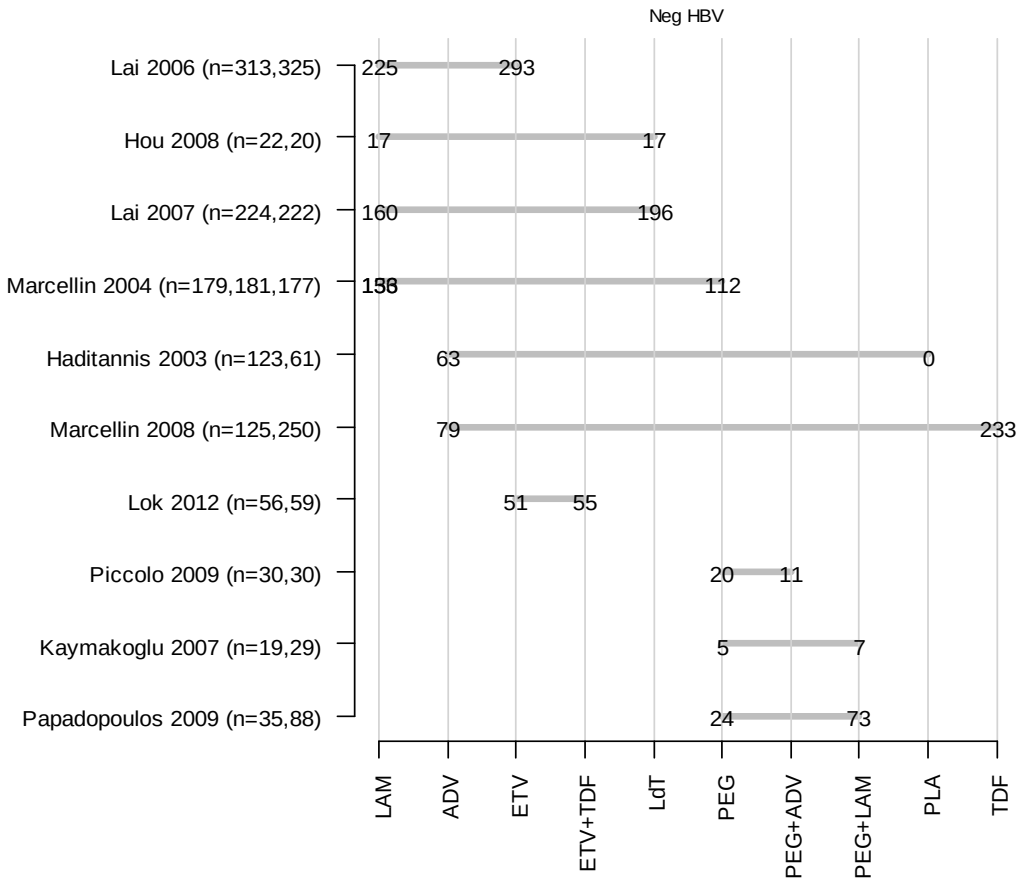
GLORIA WOO,[*,‡] GEORGE TOMLINSON,[*,‡,§] YASUNORI NISHIKAWA,[*] MATTHEW KOWGIER,[*] MORRIS SHERMAN,[‡,§] DAVID K. H. WONG,[‡,§] BA PHAM,[*,‡] WENDY J. UNGAR,[*,‡,‖] THOMAS R. EINARSON,[*,‡] E. JENNY HEATHCOTE,[‡,§] and MURRAY KRAHN[*,‡,§]

[*]Toronto Health Economics and Technology Assessment Collaborative, [‡]Departments of Medicine and Health Policy, Management and Evaluation, and Faculty of Pharmacy, University of Toronto, Toronto; [§]University Health Network, Toronto General Research Institute and Clinical Studies Resource Centre Toronto Western Research Institute, Toronto; [‖]Hospital for Sick Children Research Institute, Toronto, Ontario, Canada

Beware disconnected networks



Also beware zero event counts



Simultaneous comparison of multiple treatments: combining direct and indirect evidence

Deborah M Caldwell, A E Ades, J P T Higgins

How can policy makers decide which of five treatments is the best? Standard meta-analysis provides little help but evidence based decisions are possible

Several possible treatments are often available to treat patients with the same condition. Decisions about optimal care, and the clinical practice guidelines that inform these decisions, rely on evidence based evaluation of the different treatment options.[1 2] Systematic reviews and meta-analyses of randomised controlled trials are the main sources of evidence. However, most systematic reviews focus on pair-wise, direct comparisons of treatments (often with the comparator being a placebo or control group), which can make it difficult to determine the best treatment. In the absence of a collection of large, high quality, randomised trials comparing all eligible treatments (which is invariably the situation), we have to rely on indirect comparisons of multiple treatments. For example, an indirect estimate of the benefit of A over B can be obtained by comparing trials of A *v* C with trials of B *v* C,[3–5] even though indirect comparisons produce relatively imprecise estimates.[6] We describe comparisons of three or more treatments, based on pair-wise or multi-arm comparative studies, as a multiple treatment comparison evidence structure.

WILLIAM DEN MARK/SPL

Angioplasty balloon device used to unblock and widen arteries

Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol BS8 2PR

Deborah M Caldwell
research associate

A E Ades
professor of public health science

Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR

J P T Higgins
senior research associate

Correspondence to:
D M Caldwell
d.m.caldwell@bristol.ac.uk

BMJ 2005;331:897–900

The need to combine direct and indirect evidence

Concerns have been expressed over the use of indirect comparisons of treatments.[4 5] The Cochrane Collaboration's guidance to authors states that indirect comparisons are not randomised, but are "observational studies across trials, and may suffer the biases of observational studies, for example confounding."[7] Some investigators believe that indirect comparisons may systematically overestimate the effects of treatments.[8] When both indirect and direct comparisons are available, it has been recommended that the two approaches be considered separately and that direct comparisons should take precedence as a basis for forming conclusions.[5 7]

Difficulties arise, however, if the direct evidence is inconclusive but the indirect evidence, either alone or in combination with the direct evidence, is not. Furthermore, this approach becomes increasingly impractical as the number of treatments increases. If five treatments have been compared with each other,

P+ *Further details of the method are on bmj.com*

Probability best – what is the problem?

Table 3 Percentage mortality at 35 days and the probability that each treatment is best (lowest mortality) in multiple treatment comparison analysis*

	Fixed effect model		Random effects model	
	35 day Mortality %	Probability best	35 day Mortality %	Probability best
Streptokinase	6.7	0	6.8	0
Alteplase	6.7	0	6.5	0.003
Accelerated alteplase	5.8	0	5.8	0.001
Streptokinase + alteplase	6.5	0	6.6	0.002
Reteplase	6.1	0	6.0	0.01
Tenecteplase	5.8	0.004	5.8	0.03
Percutaneous transluminal coronary angioplasty	4.4	0.995	4.3	0.95

*Absolute mortality is based on the average mortality with streptokinase in the 19 randomised controlled trials that included it (see bmj.com for further details).

Volume 12 • Number 6 • 2009
VALUE IN HEALTH

No Study Left Behind: A Network Meta-Analysis in Non–Small-Cell Lung Cancer Demonstrating the Importance of Considering All Relevant Data

Neil Hawkins, PhD, MSc,[1] David A. Scott, MA,[1] Beth S. Woods, MSc,[1] Nicholas Thatcher, PhD, MD[2]

[1]Oxford Outcomes Ltd., Oxford, UK; [2]Department of Medical Oncology, Christie Hospital NHS Foundation Trust, Manchester, UK

ABSTRACT

Objective: To demonstrate the importance of considering all relevant indirect data in a network meta-analysis of treatments for non–small-cell lung cancer (NSCLC).

Methods: A recent National Institute for Health and Clinical Excellence appraisal focussed on the indirect comparison of docetaxel with erlotinib in second-line treatment of NSCLC based on trials including a common comparator. We compared the results of this analysis to a network meta-analysis including other trials that formed a network of evidence. We also examined the importance of allowing for the correlations between the estimated treatment effects that can arise when analysing such networks.

Results: The analysis of the restricted network including only trials of docetaxel and erlotinib linked via the common placebo comparator produced an estimated mean hazard ratio (HR) for erlotinib compared with docetaxel of 1.55 (95% confidence interval [CI] 0.72–2.97). In contrast, the network meta-analysis produced an estimated HR for erlotinib compared with docetaxel of 0.83 (95% CI 0.65–1.06). Analyzing the wider network improved the precision of estimated treatment effects, altered their rankings and also allowed further treatments to be compared. Some of the estimated treatment effects from the wider network were highly correlated.

Conclusions: This empirical example shows the importance of considering all potentially relevant data when comparing treatments. Care should therefore be taken to consider all relevant information, including correlations induced by the network of trial data, when comparing treatments.

Keywords: indirect comparison, mixed treatment comparisons, network meta-analysis, non–small-cell lung cancer.

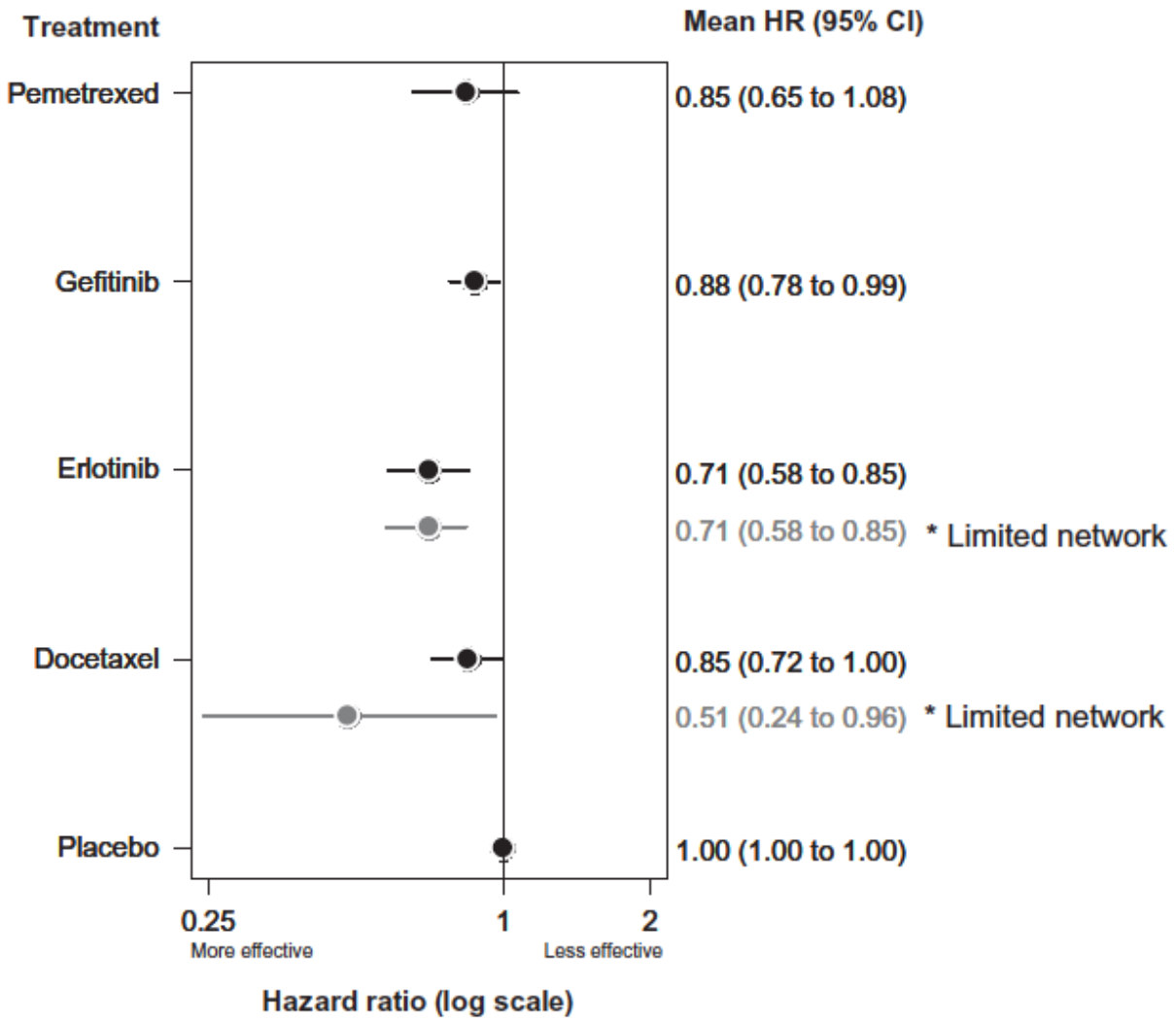
Need to show all rankings

Ranking	1	2	3	4	5
	Most effective				Least effective
Treatment					
Erlotinib	0.85	0.10	0.04	0.02	0.00
Pemetrexed	0.12	0.39	0.18	0.23	0.08
Docetaxel	0.03	0.34	0.47	0.14	0.01
Gefitinib	0.00	0.16	0.30	0.52	0.01
Placebo	0.00	0.00	0.01	0.09	0.90

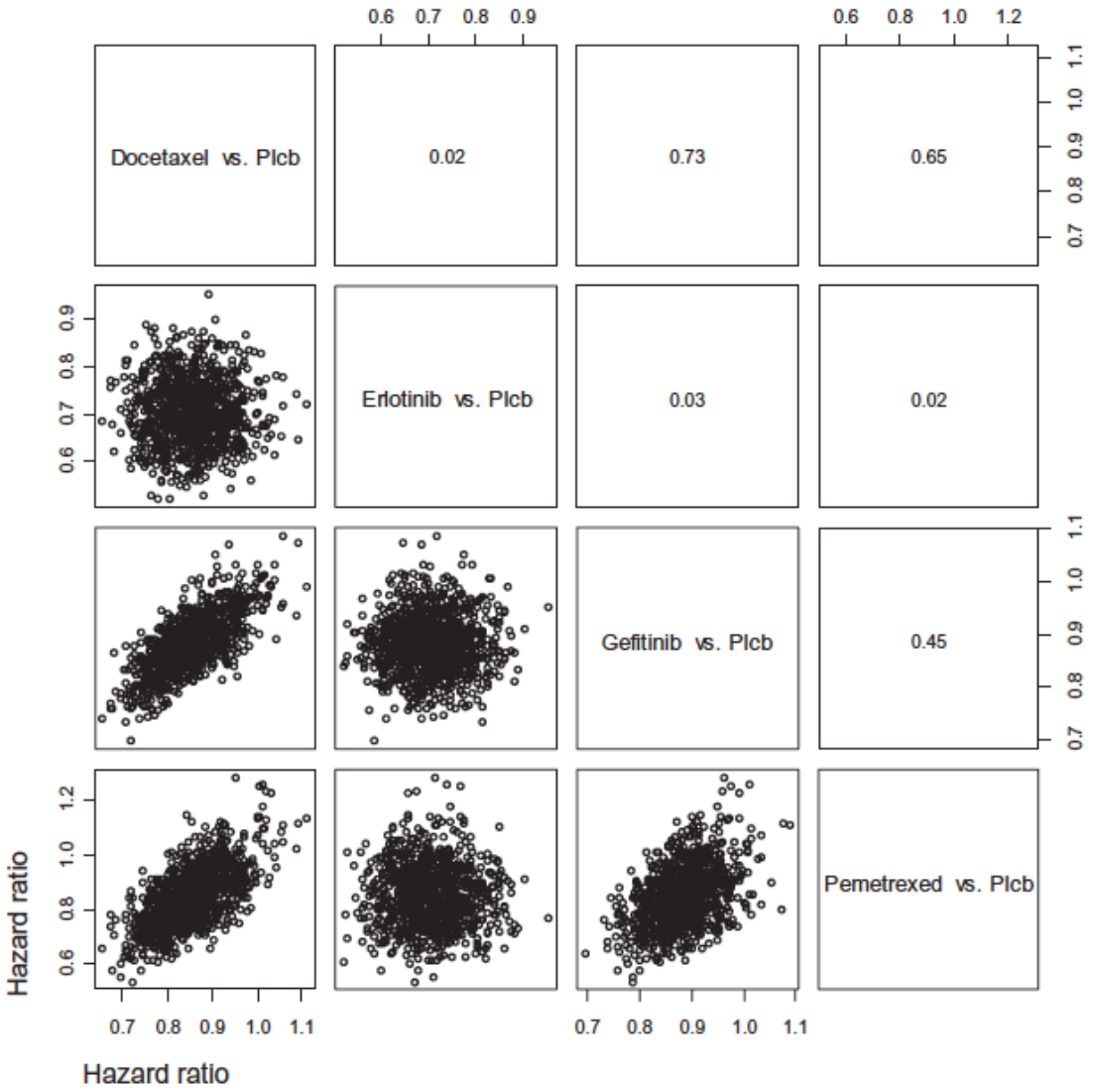
Figure 4 Network meta-analysis results displayed as probability of treatment occupying different rankings.

Does not allow for structural uncertainty ('best' case estimate)

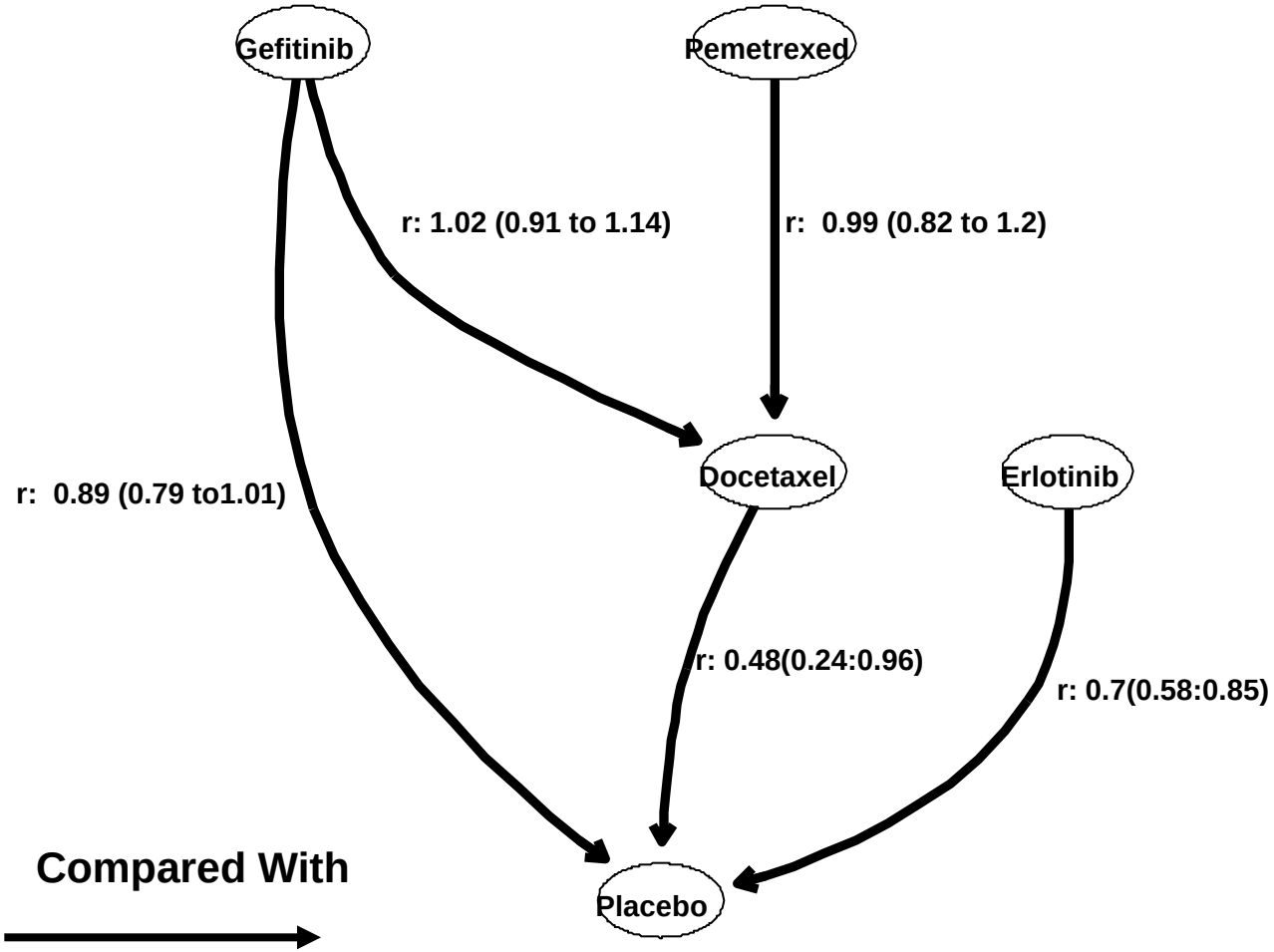
Why not use ‘forest’ plot



Because of correlations in effect estimates



Which arise from the network structure



Need to report all pairwise comparisons

	Placebo	Erlotinib	Pemetrexed	Docetaxel	Gefetinib
Placebo	-	1.43 (1.18:1.72)	1.2 (0.93:1.53)	1.18 (1:1.39)	1.14 (1.01:1.29)
Erlotinib	0.71 (0.58:0.85)	-	0.84 (0.61:1.14)	0.83 (0.64:1.06)	0.81 (0.64:1)
Pemetrexed	0.85 (0.66:1.08)	1.22 (0.88:1.65)	-	1.00 (0.82:1.2)	0.97 (0.77:1.2)
Docetaxel	0.85 (0.72:1)	1.22 (0.94:1.56)	1.01 (0.83:1.22)	-	0.97 (0.86:1.08)
Gefetinib	0.88 (0.78:0.99)	1.26 (1:1.57)	1.05 (0.83:1.3)	1.03 (0.92:1.16)	-

Figures are row treatments compared with column treatments

Some alternative presentations of uncertainty

ARTICLE IN PRESS

ELSEVIER

Journal of Clinical Epidemiology ■ (2010) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial

Georgia Salanti[a,*], A.E. Ades[b], John P.A. Ioannidis[a,c,d]

[a]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece
[b]Department of Community Based Medicine, University of Bristol, Bristol, UK
[c]Department of Medicine, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center and Tufts University School of Medicine, Boston, MA, USA
[d]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Accepted 31 March 2010

Abstract

Objective: To present some simple graphical and quantitative ways to assist interpretation and improve presentation of results from multiple-treatment meta-analysis (MTM).

Study Design and Setting: We reanalyze a published network of trials comparing various antiplatelet interventions regarding the incidence of serious vascular events using Bayesian approaches for random effects MTM, and we explore the advantages and drawbacks of various traditional and new forms of quantitative displays and graphical presentations of results.

Results: We present the results under various forms, conventionally based on the mean of the distribution of the effect sizes; based on predictions; based on ranking probabilities; and finally, based on probabilities to be within an acceptable range from a reference. We show how to obtain and present results on ranking of all treatments and how to appraise the overall ranks.

Conclusions: Bayesian methodology offers a multitude of ways to present results from MTM models, as it enables a natural and easy estimation of all measures based on probabilities, ranks, or predictions. © 2010 Elsevier Inc. All rights reserved.

Keywords: Predictive intervals; Posterior probabilities; Ranking; Network meta-analysis; Mixed-treatment comparison; Bayesian meta-analysis

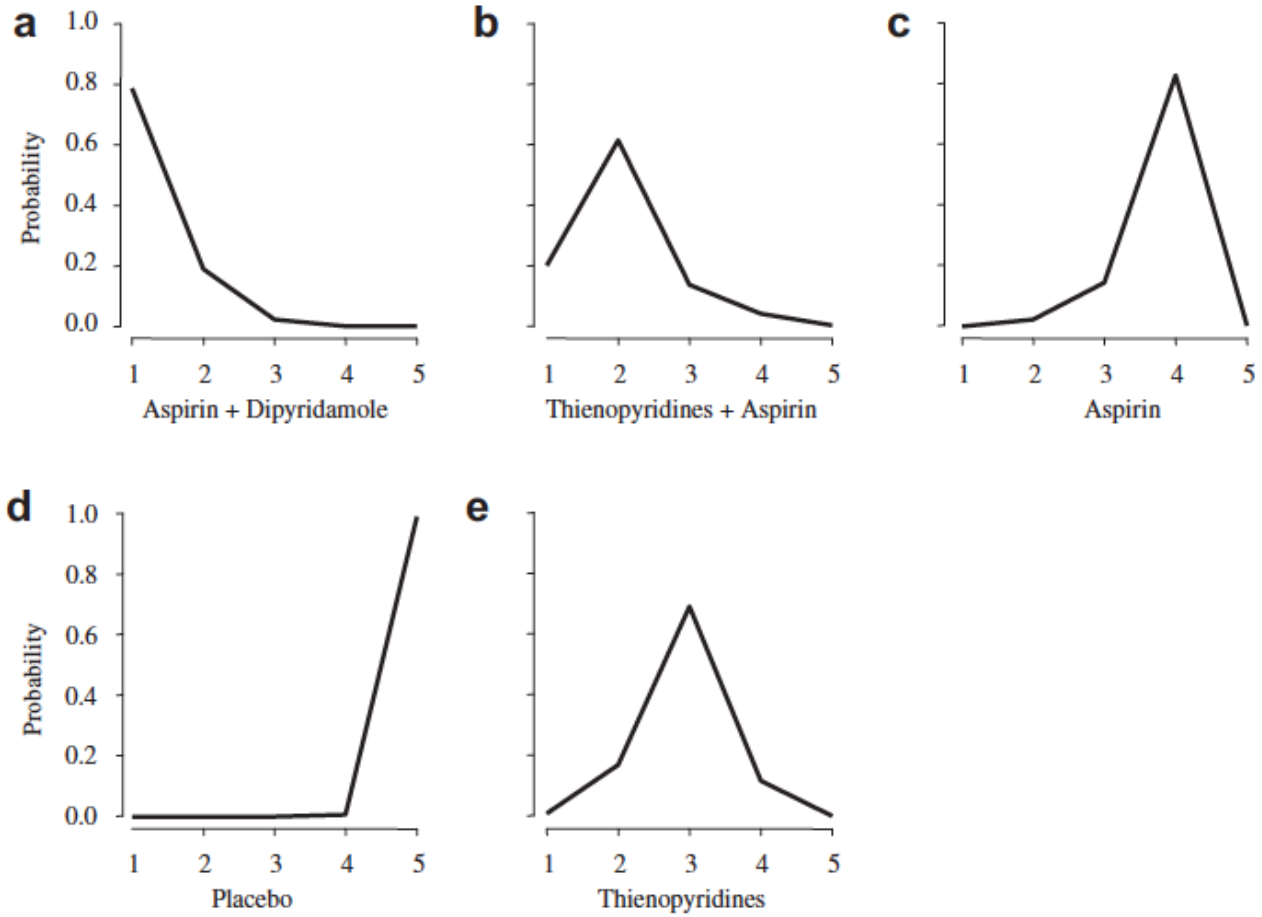


Fig. 3. Rankograms for the five antiplatelet regimens. On the horizontal axis are the five possible ranks and on the vertical axis the probability of a treatment to achieve each rank.

Table 3
Ranking of competing antiplatelet treatments

Treatment	SUCRA (%)	Median rank (95% credible interval)	Probability to be no worse than 1.1-fold
Aspirin+dipyridamole	94	1 (1, 2)	Reference
Thienopyridines+aspirin	74	2 (1, 3)	51
Thienopyridines	52	3 (2, 4)	18
Aspirin	30	4 (3, 4)	2
Placebo	0	5 (5, 5)	0

SUCRA values, median ranks (with 95% credible intervals), and probability for each treatment to increase the odds of the outcome no more than 10% compared with the best option.

Abbreviation: SUCRA, surface under the cumulative ranking curve.

Finally: from the ‘Professional Meta-Analyst’

Three aims of meta-analysis:

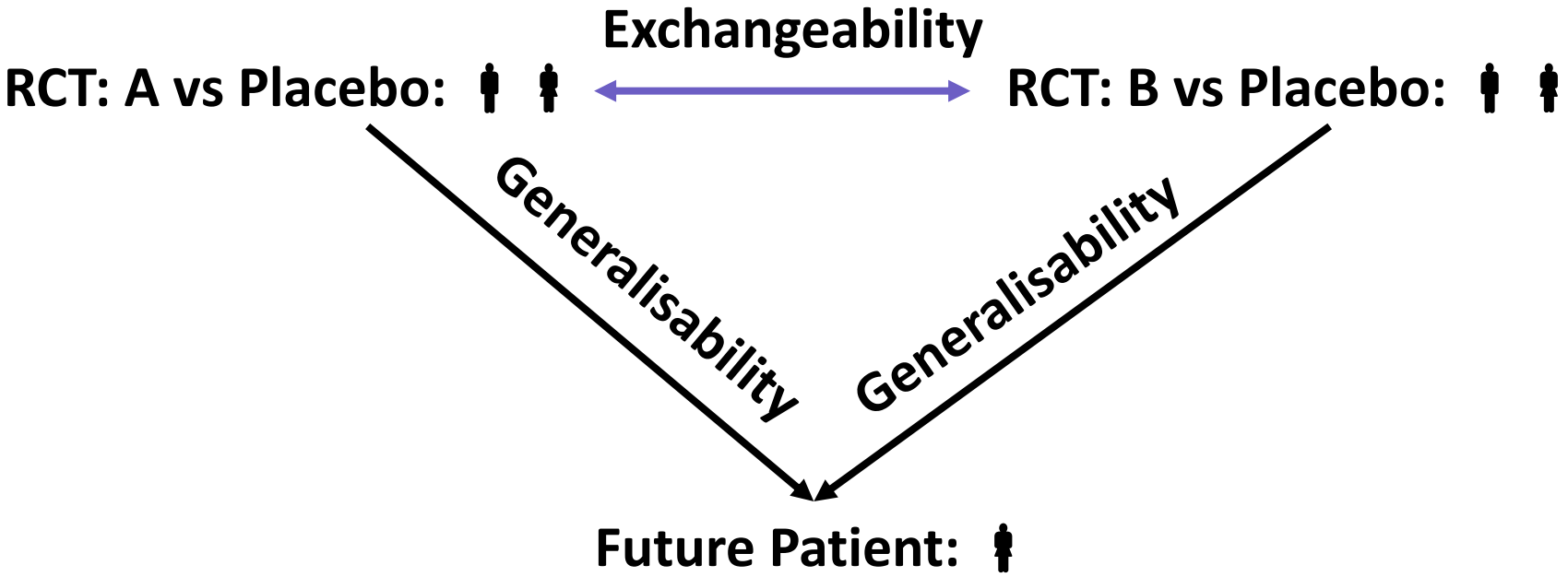
1st to obtain increased power

2nd to obtain the best risk estimate from many, often conflicting or even bewildering, studies. In its best form, it is an attempt to clarify some of the heterogeneity between studies by subgroup analysis.

3rd to answer a question which the original studies were not aimed at

Rosendall 1994. J Clin Epidemiol Vol. 47. No. 12,p p. 1325-1326

Is exchangeability implicit in clinical decision-making?



CONCLUSIONS & WRAP-UP

Conclusions

- NMA are a key component of drug development plans and support defining product “value”
- NMA enable indirect comparisons to be made with other therapies used in clinical practice but not compared in head to head randomized controlled trials
- NMA are observational with strong assumptions and need to be interpreted with caution with key limitations and biases fully described
- NMA require cross-functional engagement and alignment between clinicians, statisticians, and health economists

Recommendations

- Ensure global product development plans include NMA activities
- Plan to conduct NMAs during phase 2/3 to understand evolving clinical evidence
- Educate the fundamentals of NMAs to cross-functional partners
- Statisticians are responsible for conducting NMAs

526

Statisticians play an important role in NMAs

- Statisticians bring strategic contributions to product teams in planning NMAs
- Statisticians can plan the detailed analyses required for NMAs
- Statisticians have the technical expertise and tools to conduct extensive and robust NMAs
- Statisticians can present and appropriately interpret the results of NMAs

527

Lessons learned from case studies

- Early team input and buy in is essential
- Obtain draft data and get analysis programs in place off critical path
- Perform validation of data extraction from systematic review
- May require a large number of analyses
  - # endpoints x # treatments x # analysis sets
- Automate indirect (and mixed treatment) comparisons within SAS v9.2 (9.3)
  - use WinBUGS as a validation tool
- Publication planning

REFERENCES

References

- 2006 CADTH Guidelines for the Economic Evaluation of Health Technologies: Canada. (www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf).
- 2008 National Institute for Health and Clinical Excellence (NICE) Guide to the Methods of Technology Appraisal. (www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf)
- Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu G. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics (2006) 24:1-19.
- Agency for Healthcare Research and Quality. Methods reference guide for effectiveness and comparative effectiveness reviews. Version 1.0 Draft posted Oct 2007. (www.effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf)
- Bucher HC, Guytt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J.Clin.Epi (1997) 60(6):683-91.
- Cooper KL et al. Immune Thrombocytopenia. *Int J Technol Assess Health Care.* 2012;28:249–258
- Cochrane (2012) : Bisphosphonates in multiple myeloma: a network metaanalysis (Review)
- Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009 Feb 28;**373(9665):746-758**
- De Hoog, F. R., Speed, T. P., and Williams, E. R. (1990). On a matrix identity associated with generalized least squares. *Linear Algebra and its Applications 127, 449–456.*
- Dias S, Welton NJ, Caldwell DM and Ades. Checking consistency in mixed treatment comparison meta-analysis. AE. Stats in Med (2010) 29:932-944.
- Dias S, Welton NJ, Sutton AJ and Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials.
- Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G and Ades AE. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials.
  NICE DSU Report (2011) (www.nicedsu.org.uk)
- Eichler et al. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nature Reviews Drug Discovery* (2010); **9**, 277-291 doi:10.1038/nrd3079
- EUnetHTA guideline comparators and comparisons: direct and indirect comparisons. EUnetHTA website. Feb 2013. http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/Direct%20and%20indirect%20comparisons.pdf
- Freemantle et al. Osteoporosis, International Osteoporosis Foundation and National Osteoporosis Foundation. 2012
- Higgins & Whitehead (1996) *Borrowing strength from external trials in meta-analysis*. Statistics in Medicine, 15: 2733–2749.

References (cont.)

- ICWG Report of the Indirect Comparisons working Group to the Pharmaceutical Benefits Advisory Committee: Assessing indirect comparisons. (www.health.gov.au/internet/main/publishing.nsf/Content/B11E8EF19B358E39CA25754B000A9C07/$File/ICWG%20Report%20FINAL2.pdf)
- Institute for Quality and Efficiency in Healthcare (IQWIG). General Methods, Version 4. 2011. https://www.iqwig.de/en/methods/methods_papers/general_methods.3020.html
- Jansen J et al. Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. Value in Health 14 (2011) 417-428.
- Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC et al. Statistical approaches for conducting network meta-analysis in drug development, Pharmaceutical Statistics 2011, 10, 523-531
- Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? Simulations study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stats in Med (2005)24:2401-2428.
- Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stats in Med (2004) 23:3105-3124
- Lu G, Ades AE.. Modelling between-trial variance structure in mixed treatment comparisons. Biostatistics (2009) 10:4:792-805.
- Lumley T. network meta-analysis for indirect treatment comparisons. Stats in Med. (2002) 21:2313-2324.
- Mandema JW et al. A Dose–Response Meta-Analysis for Quantifying Relative Efficacy of Biologics in Rheumatoid Arthritis. Nature 2011. Volume 90 Number 6; 828-835.
- NICE Technical Series Documents 2011: http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm
- Pagliaro et al (1992). Prevention of First Bleeding in Cirrhosis: A Meta-Analysis of Randomized Trials of Nonsurgical Treatment . Annals of internal Medicine,117,59-70
- Piepho H. P., Williams E. R., and Madden L. V.. 2012. *The Use of Two-Way Linear Mixed Models in Multitreatment Meta-Analysis.* Biometrics.
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010, 11:32. (24 March 2010)
- Stevens, J. 2013. *A note on dealing with missing standard errors in meta-analyses of continuous outcome measures in WinBUGS.* Pharmaceut. Statist. 2011, 10 374–378
- Welton NJ, Cooper NJ Ades AE, Lu G, Sutton AJ. Mixed treatment comparison with multiple outcomes reported inconsistently across trial: evaluation of antivirals for treatment of influenza A and B. Stats in Med (2008) 27:5620-5639.
- Whitehead (2002) *Meta-analysis of controlled clinical trials*. Wiley.
- Woods ES, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. BioMed Central Medical Research methodology (2010) 10-54.