Bayesian approaches to subgroup analysis, selection problems and signal detection in drug development

David Ohlssen

Statistical Methodology, Novartis Pharmaceutical corporation

May 22nd 2013

NOVARTIS

Introduction

Many good reasons for using Bayesian methods in drug development

- Good decision making should be based on all relevant information
  - Therefore, formally accounting for contextual information makes sense
  - However, this is easier said than done
- Bayesian metrics can add value (e.g posterior probability, predictive probability)
- Bayesian approach is “easier“ in complex settings with various sources of uncertainty.

Bayesian methods applied at Novartis

A long history of using Bayesian methods

Appl. Statist. (1986),
35, *No.* 2, *pp.* 93–150

Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry

By A. RACINE, A. P. GRIEVE and H. FLÜHLER

CIBA-GEIGY AG, Basle, Switzerland

A. F. M. SMITH†

The University of Nottingham, UK

[*Read before the* Royal Statistical Society, *at a meeting organized by* the Research Section *on Wednesday, March 5th, 1986,* The President, Dr. J. A. Nelder, *in the Chair*]

SUMMARY

Four typical applications of Bayesian methods in pharmaceutical research are outlined. The implications of the use of such methods are discussed, and comparisons with traditional methodologies are given.

Keywords: Bayesian Analysis; Acute Toxicity; Probit Model; Prior Information; Clinical Trials; Two-Period Crossover Design; Carryover-Effect; Bayes factor; Bioequivalence Assessment; Historical Information; Two-Stage Procedure; Pharmacokinetics; Population Modelling; Hierarchical Model; EM Algorithm; Prediction

- Using historical data from previous studies to form priors
- Bayesian Adaptive designs in phase I Oncology
- Quantitative Decision making techniques
- Evidence synthesis
- Exploratory sub-group analysis
- Sensitivity analysis plans for handling missing data

NOVARTIS

Still many challenges moving Bayes into practice

- Some colleagues have limited formal education in Bayesian methods (varies considerably across different sites)
- Even colleagues with a good background in Bayesian statistics find it difficult to connect with practice
- Bayesian methods usually require a much greater level of engagement and resource
- Skepticism on whether Bayesian approaches really add value

DIA Bayesian Scientific Working Group

Group of representatives from Regulatory, Academia, and Industry, engaging in scientific discussion/collaboration

- facilitate appropriate use of the Bayesian approach
- contribute to progress of Bayesian methodology throughout medical product development

NOVARTIS

Vision

Ensure that Bayesian methods are

well-understood, accepted, and broadly utilized for design, analysis, and interpretation to **improve patient outcomes**

throughout the medical product development process and to improve decision making.

Part 1 Motivating examples

subgroup analysis, selection problems and signal detection

Challenges with exploratory subgroup analysis

random high bias - Fleming 2010

Effects of 5-Fluorouracil Plus Levamisole on Patient
Survival Presented Overall and Within Subgroups, by Sex and Age*

Hazard Ratio Risk of Mortality

Analysis Group	North Central Treatment Group Study (n = 162)	Intergroup Study # 0035 (n = 619)
All patients	0.72	0.67
Female	**0.57**	**0.85**
Male	**0.91**	**0.50**
Young	0.60	0.77
Old	0.87	0.59

Assessing treatment effect heterogeneity in multi-regional clinical trials

- Multiregional trials popularized by the need to enroll a large number of patients in a timely manner
- Interest in the consistency of treatment effects across regions (ICH E5, PMDA guidelines)
- Example - Large cardiovascular outcomes trial known as ‘PLATO’, where substantial evidence of regional heterogeneity emerged during the analysis

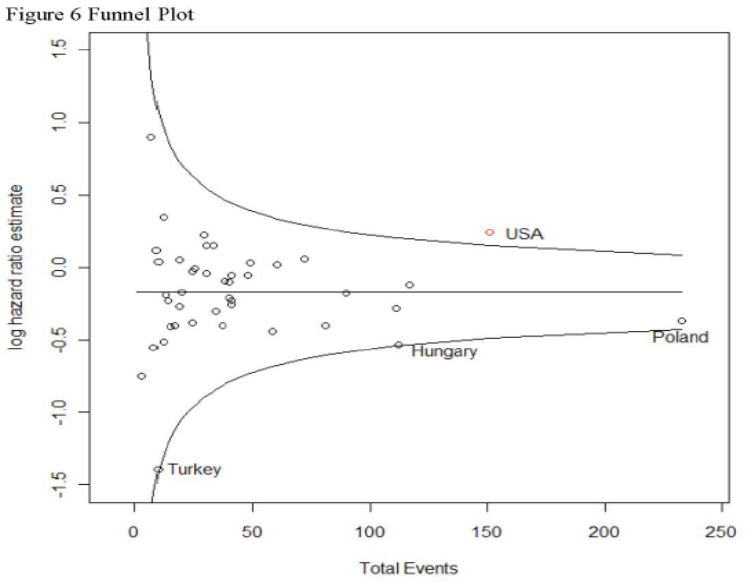
PLATO trial example

- Randomized double-blind study comparing ticagrelor (N=9333) to clopidogrel (N=9291), both given in combination with aspirin, in patients with acute coronary syndromes.
- Primary endpoint was time to first occurrence of CV death, MI or stroke.
- Randomization across 41 countries.
- Primary endpoint met for ticagrelor 9.8% vs 11.7% events HR = 0·84 95% CI 0·77–0·92]; p=0·0003.

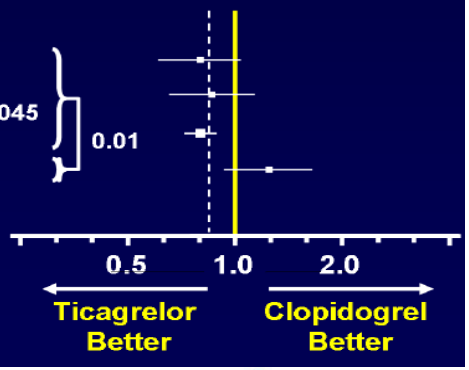
Part of the pre-specified subgroup analysis

Extracted from the FDA advisory committee material

- 31 pre-specified subgroup tests
- No adjustment for multiplicity
- Indication of variability between regions
- North America results driven by US (HR=1.27 0.92,1.75)

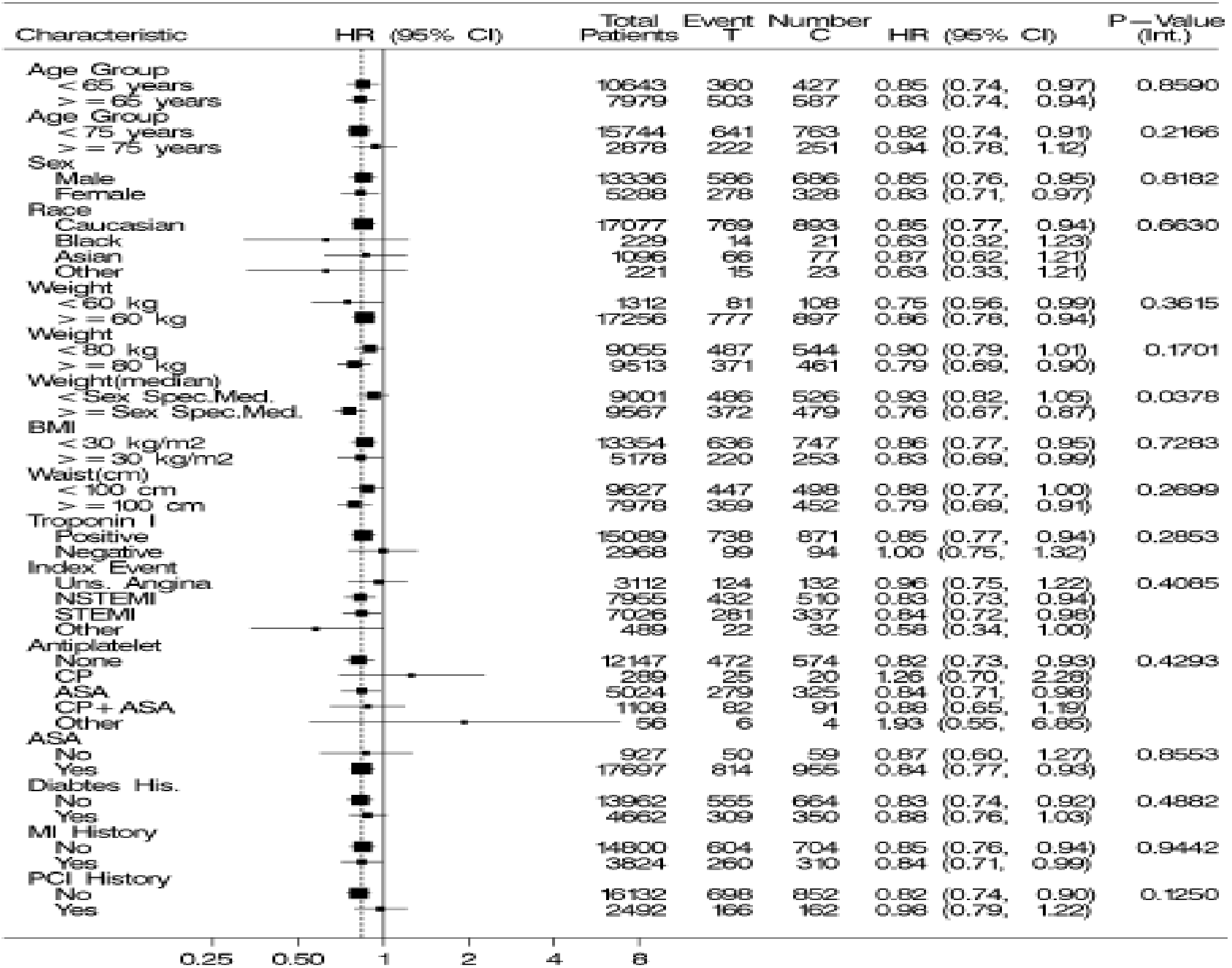


Characteristic	Total Patients	# of Events Tic	# of Events Clop	HR (95% CI)	Interaction p-value
Geographic Region					
Asia / Australia	1714	95	116	0.80 (0.61, 1.04)	0.045
Cent / Sth America	1237	91	104	0.86 (0.65, 1.13)	
Euro / Md E / Afr	13859	576	712	0.80 (0.72, 0.90)	
North America	1814	102	82	1.25 (0.93, 1.67)	0.01



NOVARTIS

Figure 13 **Hazard ratios and rates of the primary clinical endpoint by patient subgroups (PLATO - full analysis set)**

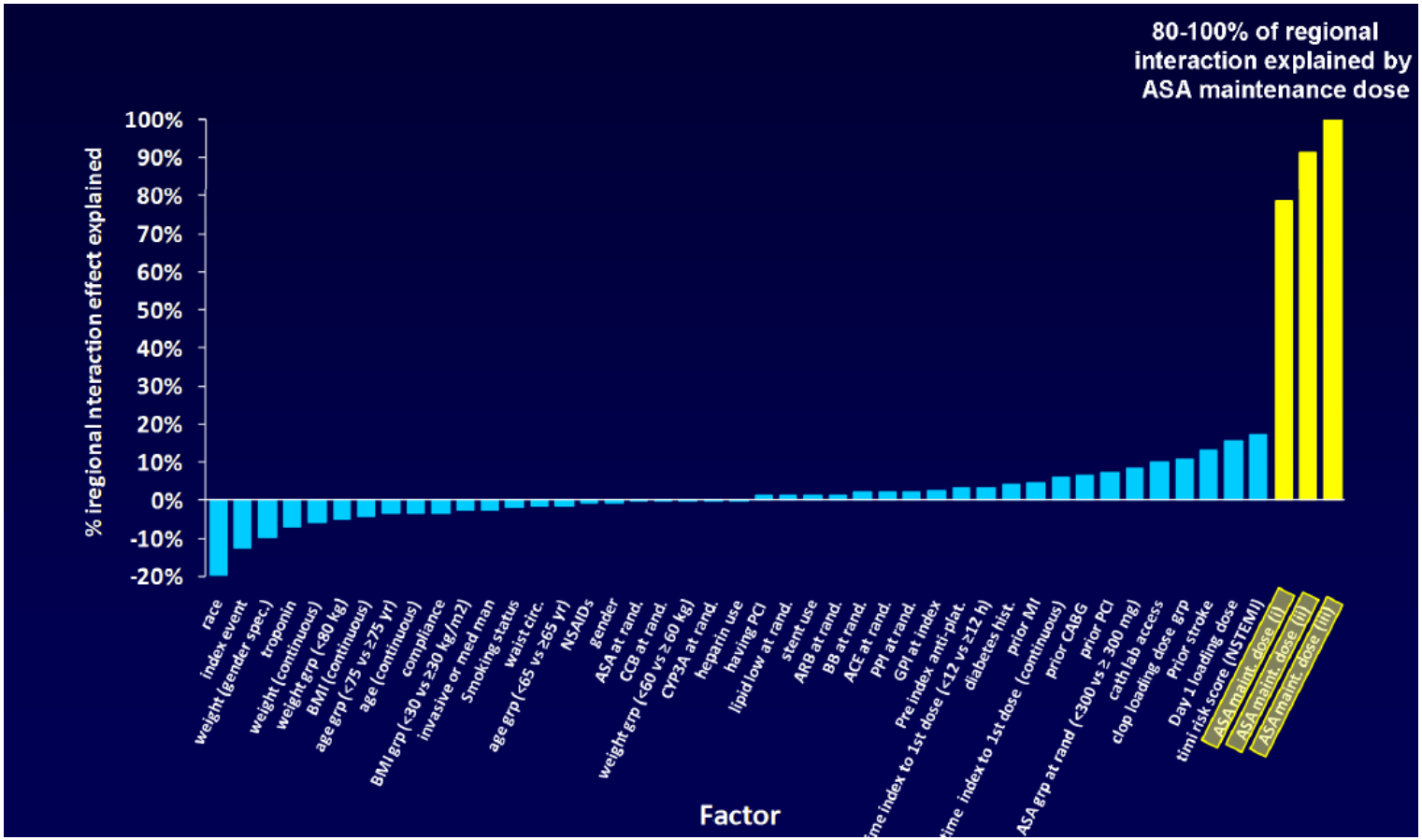


Possible explanations given in the AZ briefing material

- Errors in study conduct
  - Ruled out
- Chance
  - probability of observing a result that numerically favors clopidogrel in at least 1 region is 28% and the probability of observing a result numerically favoring clopidogrel in the NA region while numerically favoring ticagrelor in the other 3 regions is 10%.
  - FDA: chance cannot be ruled out but interaction with US/non-US is both striking and worrying
- Imbalances between US and non-US populations in patient characteristics, prognosis, or clinical management resulting in differential outcomes.

Aspirin dose a possible explanation

Astra Zeneca put forward the case that the difference between Aspirin dose when comparing US to non-US was a possible cause



Extracted from the AZ core slides used at the 2010 Advisory committee

NOVARTIS

Advisory committee vote and FDA decision memo

- The Ticagrelor NDA was presented to the Cardio-Renal Advisory committee. By a 7 to 1 vote they recommended approval
- “Although I consider the likelihood that the US/OUS was chance, a credible basis for approval for ticagrelor, I believe the evidence that aspirin dose explains the difference is a powerful further basis for approval...”
- “Labeling will note in several places, including Boxed Warning, that ticagrelor has been studied in combination with aspirin and doses above 100 mg appear to decrease effectiveness”

Some additional notes from Carroll and Fleming (2013)

- Trials are seldom powered to address pre-specified hypotheses about regional interactions.
- Such interactions usually are assessed in an exploratory manner, often with many other supportive analyses.
- As such, the first step in examining an apparent regional interaction is to assess the likelihood it is due to chance. This might include:
  - A Galbraith plot for effects within regions, and again for effects within country if possible.
  - **Bayesian subset analyses** and shrinkage estimators of regional effects
  - Lastly, replication of an observed regional interaction in a second, independent trial should be sought where possible.

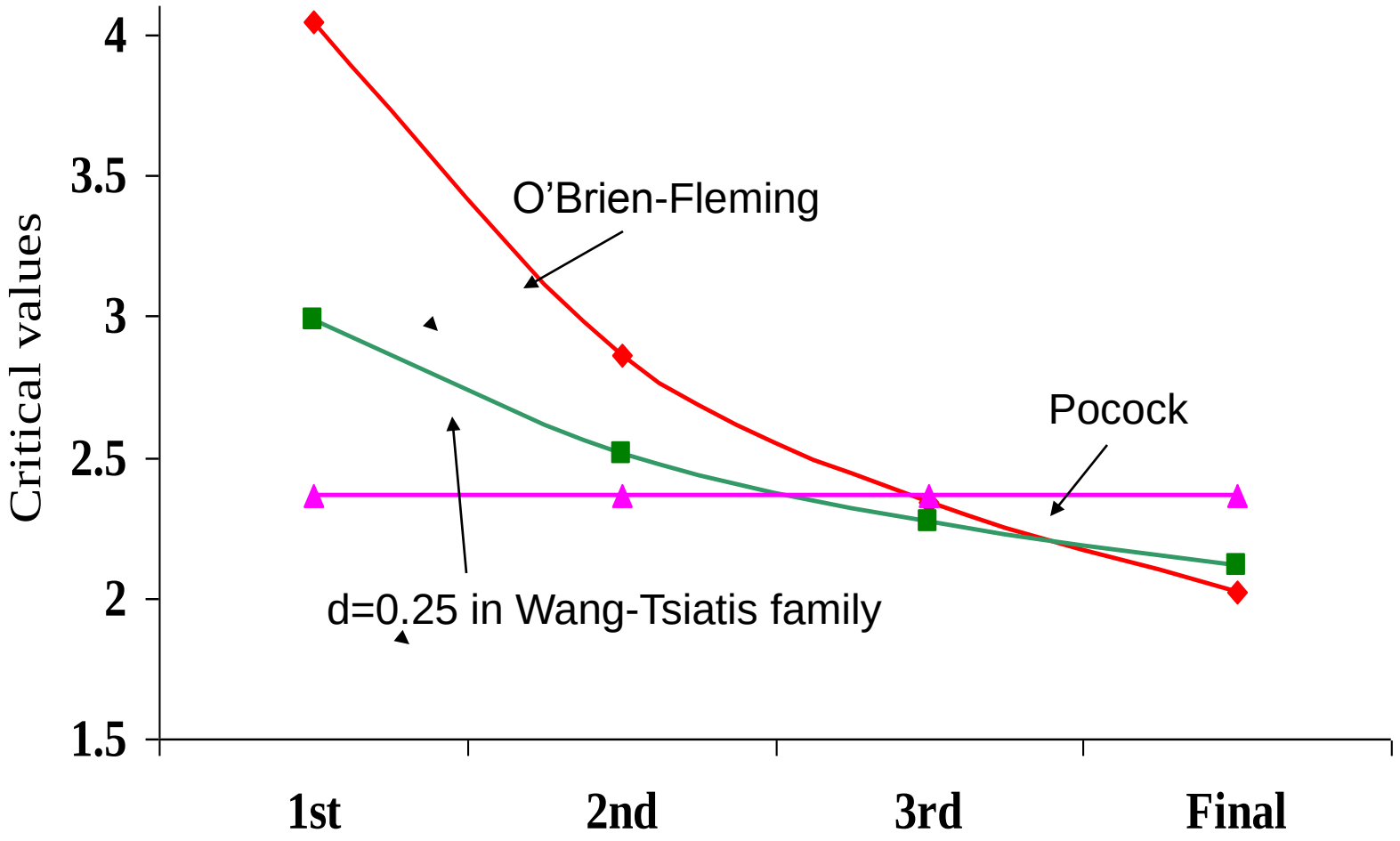
Classical group sequential design

- A framework that allows k chances to stop for success with type one error control
- More formally, we have to find critical values $z_1, z_2, \ldots, z_k$ as a solution of the integral:

$$P(Z_1 < z_1, Z_2 < z_2, \ldots, Z_k < z_k \mid H_0) = 0.975$$

- with the correlation structure of the MVN distribution determined by the amounts of data available at the analyses
- Group sequential methodology essentially boils down to imposing enough structure or constraints to determine solutions.

Example: superiority boundaries – 4 looks



NOVARTIS

Over-estimation in group sequential designs

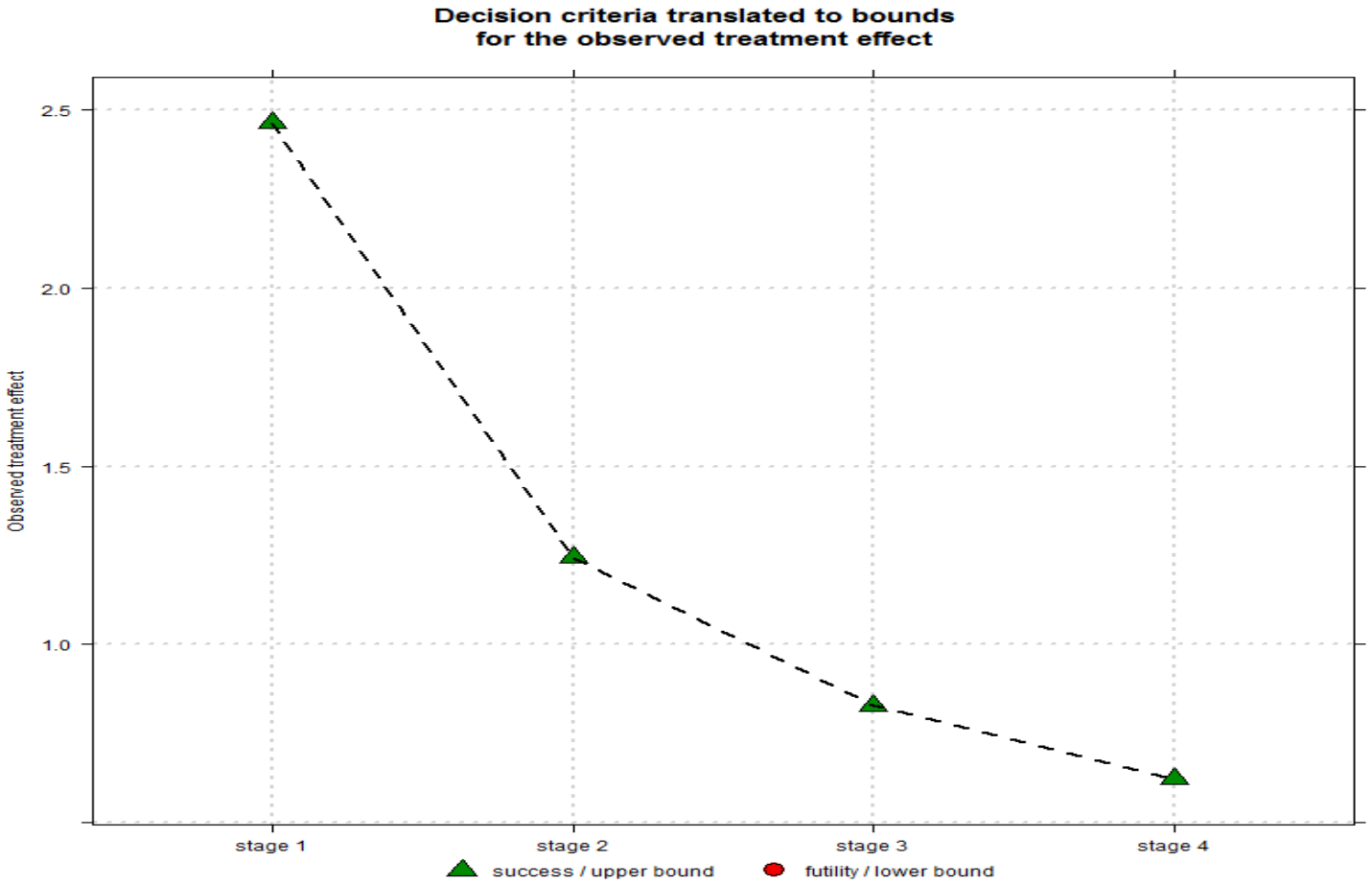
- **Overestimation in GSDs**

"…a trial terminated early for benefit will tend to overestimate true effect; this happens because there always is variability in estimation of true effect, and when assessing data over time, evidence of extreme benefit is more likely obtained at times when the data provide a random overestimate of truth."

Ellenberg, DeMets, and Fleming JAMA, 2010

O'Brien-Fleming rule on the treatment effect scale

Sd=2.17 n=100 per group



Assumed treatment effect=1

NOVARTIS

Bayesian group sequential designs

- When presenting a final treatment effect prior information could be utilized to shrink towards the hypothesized treatment effect (see Pocock and Hughes; 1990)
- Spiegelhalter et al. (2004) showed a more traditional sceptical prior centered at the null or 0 treatment effect can also be used
  - For four equally spaced IA a sceptical prior with 0.25 of the total sample size could be used leading to type one error control with a Bayesian decision rule and automatic shrinkage
  - i.e. If the Bayesian decision rule $Pr(\delta > 0|Data) > 0.975$ then the probability of achieving this under the null is 0.025.

R package available for design investigation

Package 'gsbDesign'

May 8, 2012

Type Package

Title Group Sequential Bayes Design

Version 0.95

Date 2012-05-07

Author Florian Gerber, Thomas Gsponer

Depends gsDesign, lattice, grid

Maintainer <flora.fauna.gerber@gmail.com>

Description Group Sequential Operating Characteristics for Clinical,Bayesian two-arm Trials with known Sigma and Normal Endpoints.

Safety signals

- **Rofecoxib (Vioxx, Merck)**
  - was withdrawn in 2004 due to increased risk of cardiovascular disease in patients taking drug for more than 18 months
  - Jüni et al. (2004) claimed drug should have been withdrawn several years earlier

Rofecoxib (Vioxx)

- Following the APPROVe study (Bresalier et al, *NEJM,* 2005) Rofecoxib was withdrawn in 2004 due to increased risk of cardiovascular disease in patients taking drug for more than 18 months
- Jüni et al. (2004) conducted a retrospective cumulative meta-analysis and used the results to argue the compound should have been withdrawn several years earlier

A Retrospective Cumulative meta-analysis

Rofecoxib (Vioxx) example

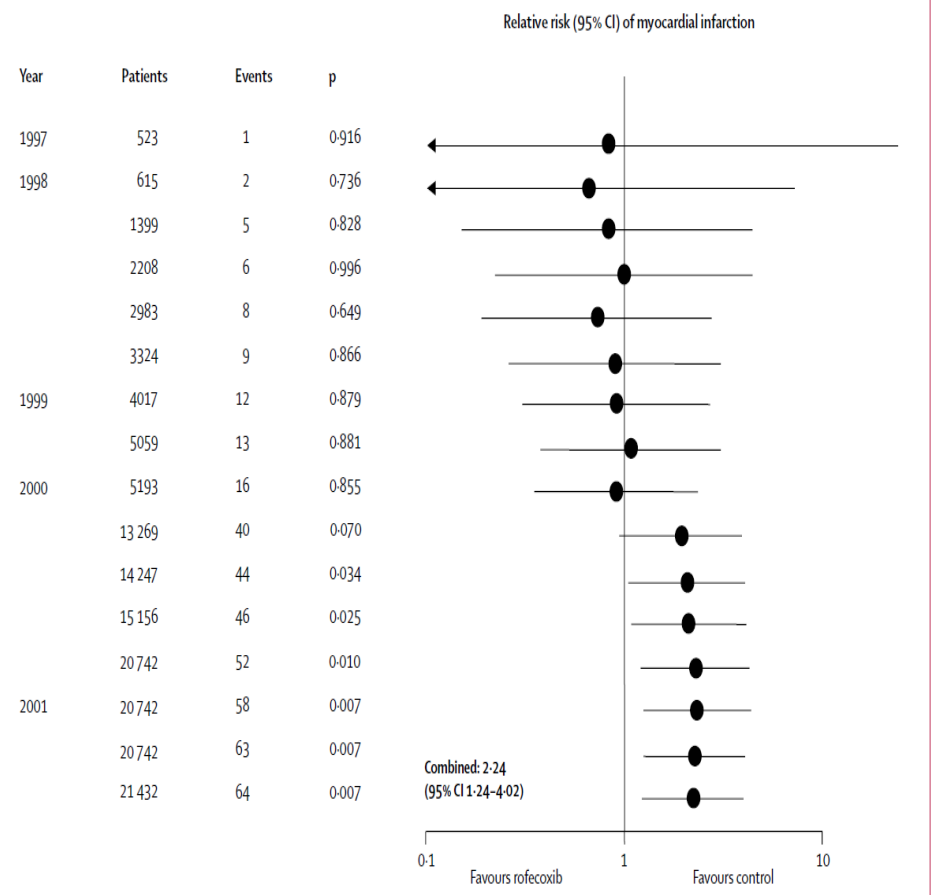


Table 1

Sequence of Studies and Comparator Usage in Juni *et al.* Figure 3

Protocol Number	Comparators	Year
029	Placebo	1997
029 extension	Diclofenac	1998
035	Diclofenac	1998
040	Placebo, Ibuprofen	1998
045	Placebo, Ibuprofen	1998
058	Placebo, Nabumetone	1998
034	Diclofenac	1999
085	Placebo, Nabumetone	1999
068 ext	**Naproxen**	2000
088, 089 (VIGOR)	**Naproxen**	2000
090	Placebo, Nabumetone	2000
096	**Placebo, Naproxen**	2000
102 (ADVANTAGE)	**Naproxen**	2000
096 ext	**Naproxen**	2001
097 ext	**Naproxen**	2001
120, 121	Placebo	2001

NOVARTIS

Discussion on the analysis of *Jüni et al*

- A careful look at the plot reveals that the large VIGOR study, designed to look at Gastro-intestinal side effects, is the most influential study in the cumulative meta-analysis
- In the VIGOR study Naproxen was the control treatment
- At the time it was argued that the imbalance in cardiovascular safety was due to the cardio-protective effect of Naproxen

Response to Jüni *et al*

- Kim and Reicin (2005) responded to Jüni et al. (2004)

"The analysis by Peter Jüni and colleagues contravenes the basic principle of meta-analyses to combine like with like, and thus arrives at flawed conclusions. "

- The concern relates to conducting a meta-analysis comparing Rofecoxhib to any control treatment rather than separate analyses for each control treatment

Discussion of Jüni *et al* example cumulative meta-analysis

- Is a basic principle of meta-analysis to combine like with like?
- It depends on the question you wish to answer
- ICH E9 suggests

"The results from trials which use a common comparator (placebo or specific active comparator) should be combined and presented separately for each comparator providing sufficient data"

- So according to ICH E9 both questions are of interest and could be examined through meta-analysis
- An alternative approach would be to use network meta-analysis, which will be discussed later in the context of Non-steroidal anti inflammatory drugs (NSAIDs) such as rofecoxhib

Overview of Bayesian techniques

Estimation or Testing?

- Is our primary purpose is to more formally **detect** unusual subgroups/ safety signals or is it to provide a **better summary** of the data and understand **treatment effect heterogeneity**?
- The question can be thought of as deciding between an estimation approach or a testing approach

Bayesian approaches to testing (1)

- Full Bayesian modeling
  - Essentially some kind of mixture model where a null distribution is included and an alternative distribution for subgroups or safety effects that are unusual
  - Calculate the posterior probability that each subgroup belongs to the alternative
  - Such posterior probabilities have the advantage that they automatically incorporate adjustments for multiple comparisons (as long as the hyper-priors are placed on the probabilities of belonging to each component of the mixture)

- Challenge Bayarri and Morales (2003) stated that

  'From a Bayesian point of view, testing whether an observation is an outlier is usually reduced to a testing problem concerning a parameter of a contaminating distribution. This requires elicitation of both (i) the contaminating distribution that generates the outlier and (ii) prior distributions on its parameters. However, very little information is typically available about how the possible outlier could have been generated.'

Bayesian testing (1) – some literature

- Berry and Berry (2004) – in the context of safety signal detection
  - Utilized shrinkage techniques and hierarchical modeling to borrow strength within and between
  - Mixture modeling to identify signals
- Sivaganesan S, Laud PW, and Müller P. (2011)
  - Subgroup analysis of clinical trial data using a zero-enriched Polya Urn
- These types of models can be quite sensitive to prior specifications so typical need simulations with frequentist operating characteristics to work out likely properties

Bayesian approaches to testing (2)

- Pragmatic Bayesian approach based on using model diagnostics
- Set-up a model that characterizes null behavior utilize Bayesian model diagnostics, typically leading to frequentist p-values to assess for outliers/ signals
- Examples - Bayarri, M. J. and Castellanos, M. E. (2007) Marshall and Spiegelhlater (2007)
- Still have the problem of dealing with multiple p-values and dependence. Could apply Bayesian FDR type methods in a second stage of analysis

Utilizing Bayesian estimation techniques

- Some examples
  - Using Bayesian hierarchical modeling, appropriate exchangeability and shrinkage to help account for reproducibility
  - Using Bayesian evidence synthesis techniques
  - Using prior structure to introduce skepticism

- Challenges
  - While these techniques can potentially help account for reproducibility they don't typically tackle multiplicity (at least directly)
  - Many possible modeling structures so how can we make sure we base conclusions on a useful model

Bayesian estimation – some literature

- Using priors
  - Pocock and Hughes (1990) - Group sequential designs
  - Simon (2002) - Bayesian subset analysis
- Hierarchical modeling
  - Dumouchel (2012) – safety example that is similar to Berry and Berry (2004) but no mixture modeling part
  - We will look at Jones et al (2011) – Exploratory subgroup analysis
- Evidence synthesis – many papers
  - We will look at Ohlssen et al (2013) – Network meta-analysis in the context of drug safety

Subgroup analysis

Acknowledgements

Hayley Jones, Beat Neuenschwander, Amy Racine, Mike Branson

Main reference

Jones, Ohlssen, Neuenschwander, Racine, Branson (2011). Bayesian models for subgroup analysis in clinical trials. *Clinical Trials* **8** 129 -143

Outline

- Introduction to subgroup analysis and Bayesian methods
- Shrinkage
- Models
- Case Study
- Concluding Remarks

Introduction to Subgroup analysis

- For biological reasons treatments may be more effective in some populations of patients than others
  - Risk factors
  - Genetic factors
  - Demographic factors
- This motivates interest in statistical methods that can explore and identify potential subgroups of interest

Introduction

Various Aspects

(Focus of this talk in **bold**)

- Definition of subgroups
  - Prospective vs. **retrospective** definition
  - **"small"** vs. very large number of subgroups

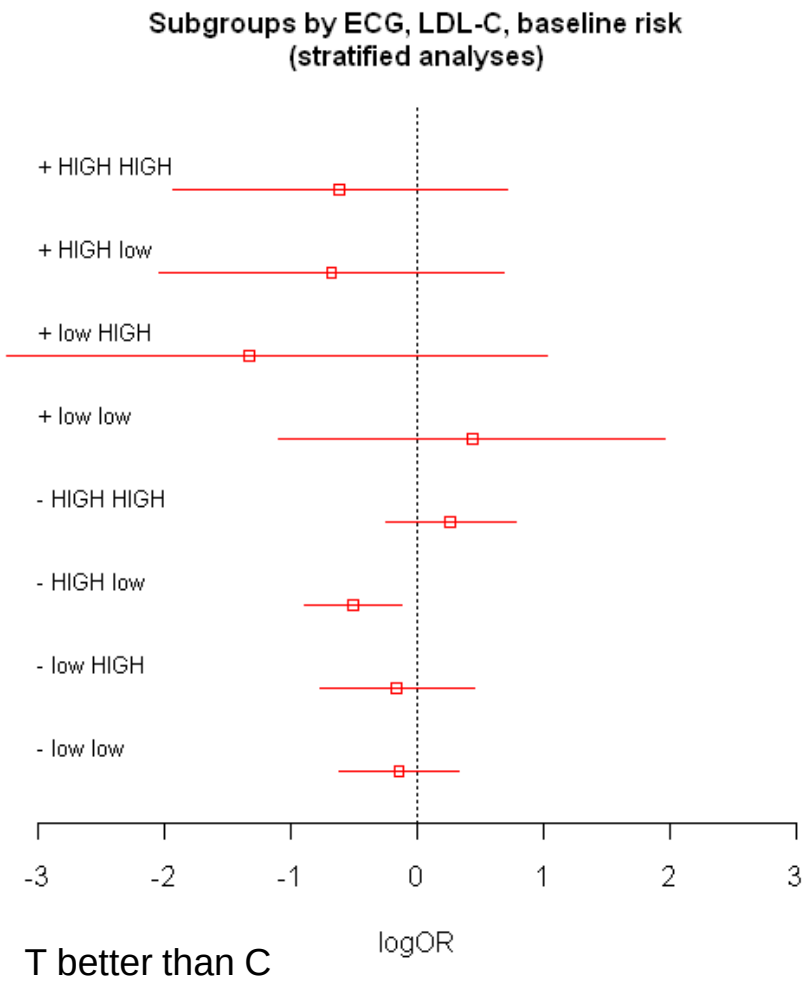
    (a few important factors that are considered predictive vs. data-mining)
- Safety vs. **efficacy**
- Testing (default "decision-making") vs. **estimation (inference)**
- **One** trial vs. **multiple** trials
- Frequentist vs. **Bayesian**
- **...**

The Bayesian modeling strategy used here

- Priors are carefully selected that we hope are dominated by the data
- Models fitted using Markov chain Monte Carlo (MCMC) estimation
- A variety of modeling structures examined
  - Model support measured using the deviance information criteria (DIC) Model diagnostics with frequentist properties used to help show whether a model has good calibration
  - Examine if similar conclusions are reached from well supported models to check inference robustness
- This work follows the ideas of Box (1980), who advocated the use of an iterative cycle of model criticism and estimation

Example 1

Data from one study

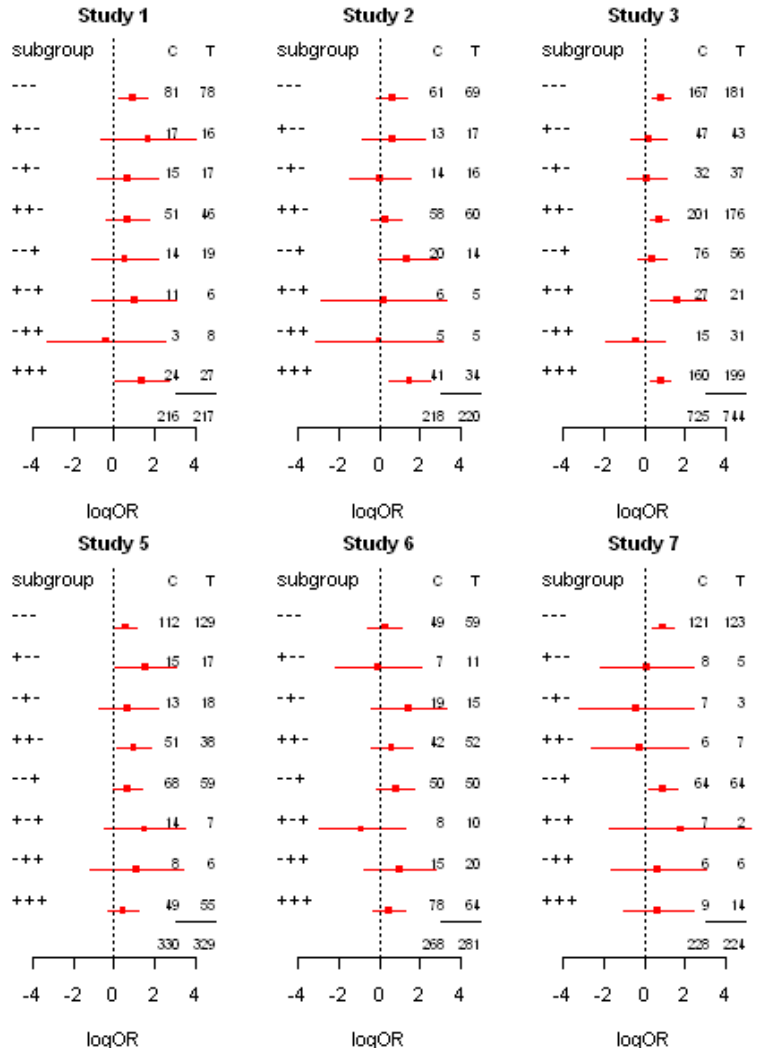


Davis & Leffingwell, Contr Clin Trials 1990)

- Endpoint
  - Coronary Heart Disease (CHD) death and Myocardial Infarction
- Comparison
  - diet + placebo (C)
  - diet + cholestyramine (T)
- Subgroups defined by baseline characteristics
  - ECG (positive/negative)
  - LDL cholesterol (high/low)
  - Risk score (including systolic blood pressure, age, smoking)

Example 2 (case study)

Data from several studies



- Subgroup analysis in a meta-analytic context
- Efficacy comparison T vs. C
- Data from 7 studies
- 8 subgroups
  - defined by 3 binary base-line covariates A, B, C
  - A, B, C high (+) or low (-)
  - describing burden of disease (BOD)
- Idea: patients with higher BOD at baseline show better efficacy

Approaches

Testing / Estimation

- Testing
  - typical for pre-planned analysis, pre-specified subgroups
- (Model-based) estimation
  - retrospective analyses

Testing Approaches

- Subgroup analysis formulated as a testing problem
  - Standard approach
    - test for treatment by subgroup interaction
    - If significant: proceed to estimate within subgroup effects
    - Pocock et al. (StatMed 2002), Assman et al. (Lancet 2000), Brookes et al. (J of Clin Epi 2004)
  - What's often done
    - Fully stratified analysis: estimates for treatment effects in each subgroup without any reference to the data in other subgroups
    - This is problematic. Berry (Biometrics 1990), Grouin et al. (JBS 2005)
  - Recommendations
    - Careful pre-planning of subgroup analysis
    - Post-hoc analyses should address the random high bias problem

Estimation Approaches

- Various approaches to estimate subgroup effects
- Instead of looking at subgroups in a fully stratified way, it is assumed that information from other subgroups carries information about subgroup(s) of interest
- Subgroup effects $\theta_1, \theta_2, \ldots, \theta_G$ are related/similar to a certain degree.
  Requirement: a reasonable assumption/model
- Under such assumptions
  - results will be different from fully stratified analysis
  - due to borrowing from the other subgroups
  - $\rightarrow$ modified point estimates
  - $\rightarrow$ generally shorter confidence intervals

Assumptions to deal with extremes

Jones et al (2011)

1) Full stratification $\theta_1, \ldots\ldots, \theta_G$

   $\Rightarrow$Assumes a different treatment effect in each subgroup

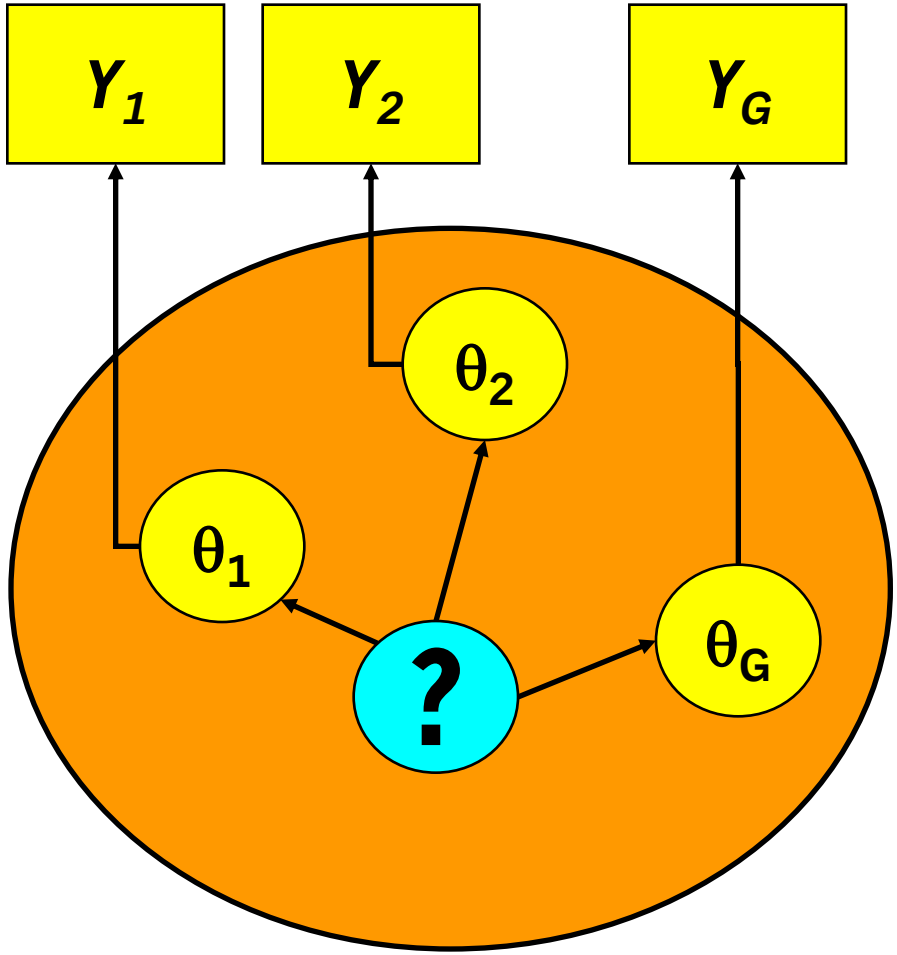
2) Equal Parameters $\theta_1 = \ldots = \theta_G$

   $\Rightarrow$ Assumes the same treatment effect in each subgroup

3) Compromise.
Effects are similar/related to a certain degree

Shrinkage estimation

Shrinkage



$Y_1,..,Y_G$
Data from *G* subgroups

$\theta_1,\ldots,\theta_G$
effects

?
Unknown 'Relationship/Similarity'

Range of possibilities:

- from same effects
- … to very different effects

Shrinkage

The simplest model

- *G* subgroups with effects $\theta_1, \theta_2, \ldots, \theta_G$
- Why shrinkage?
  - Estimates are typically more spread out than true effects $\theta_1, \theta_2, \ldots, \theta_G$
  - Extreme stratified subgroups estimates are typically too extreme
- Simple shrinkage for subgroup analyses
  - $Y_g \sim N(\theta_g, s_g^2)$, $g = 1, \ldots, G$
  - $\theta_1, \theta_2, \ldots, \theta_G \sim N(\mu, \omega^2)$
  - See Louis (JASA 1984), Davies & Leffingwell (Contr Clin Trials 1990), both using empirical Bayes techniques
- Inference
  - Classical random-effects analyses
  - Empirical Bayes
  - Fully Bayesian (with priors for μ and τ)

Fitting a standard shrinkage model when ω is unknown

- Even inference for the simple shrinkage models inference is challenging when ω is unknown
- Classical ways to address this
  - Method of moments or Mixed models framework (REML, ML GLMM)
  - Requires empirical Bayes to get at the subgroup effects
  - Difficult to account for the uncertainty surrounding ω
- Bayesian approach can be applied using MCMC estimation
  - Can be sensitive to choice of prior particularly for ω
  - Automatically propagates uncertainty surrounding ω

Shrinkage

Example 1 (Davis & Leffingwell 1990)

CHD deaths and myocardial infarction by subgroup and treatment group

	ECG	LDL.C	risk	rC	nC	rT	nT	pC	pT	logOR	logOR.se
1	+	HIGH	HIGH	7	23	5	26	30.4%	19.2%	-0.608	0.673
2	+	HIGH	low	6	32	4	38	18.8%	10.5%	-0.674	0.696
3	+	low	HIGH	3	19	1	21	15.8%	4.8%	-1.322	1.202
4	+	low	low	3	30	5	34	10%	14.7%	0.439	0.778
5	-	HIGH	HIGH	30	265	38	266	11.3%	14.3%	0.267	0.261
6	-	HIGH	low	73	665	46	664	11%	6.9%	-0.505	0.197
7	-	low	HIGH	25	268	21	260	9.3%	8.1%	-0.158	0.310
8	-	low	low	40	598	35	597	6.7%	5.9%	-0.141	0.239

$$logOR = log(rT/(nT-rT)) – log(rC/(nC-rC))$$

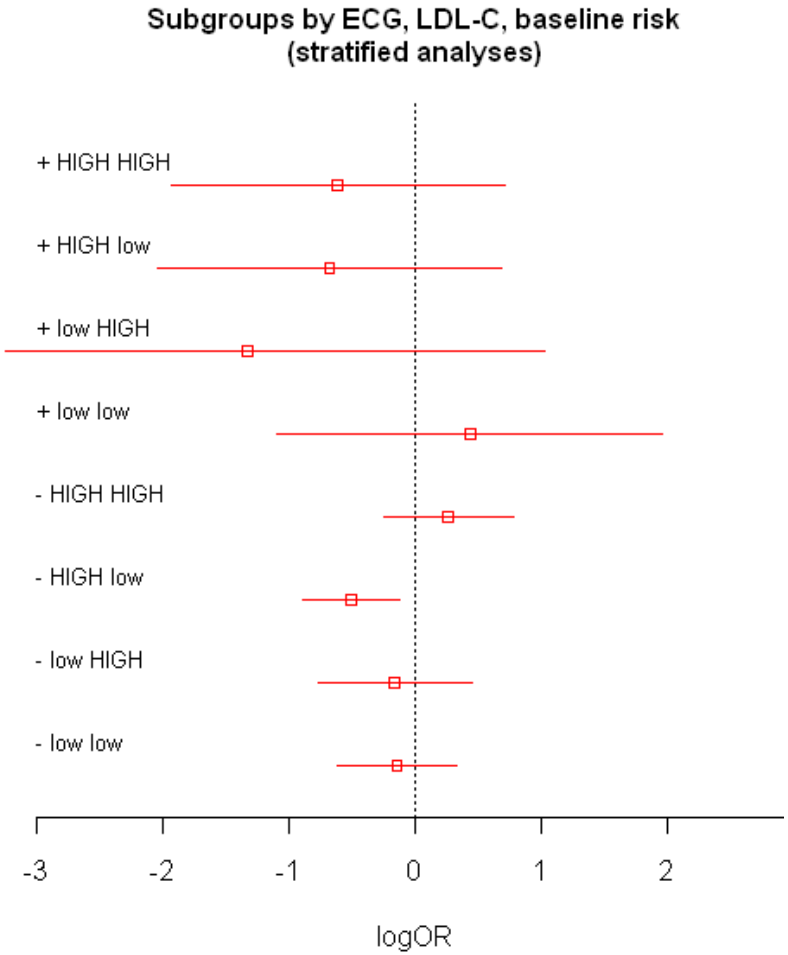
$$logOR.se = (1/rT + 1/(nT-rT) + 1/rC + 1/(nC-rC))^{1/2}$$

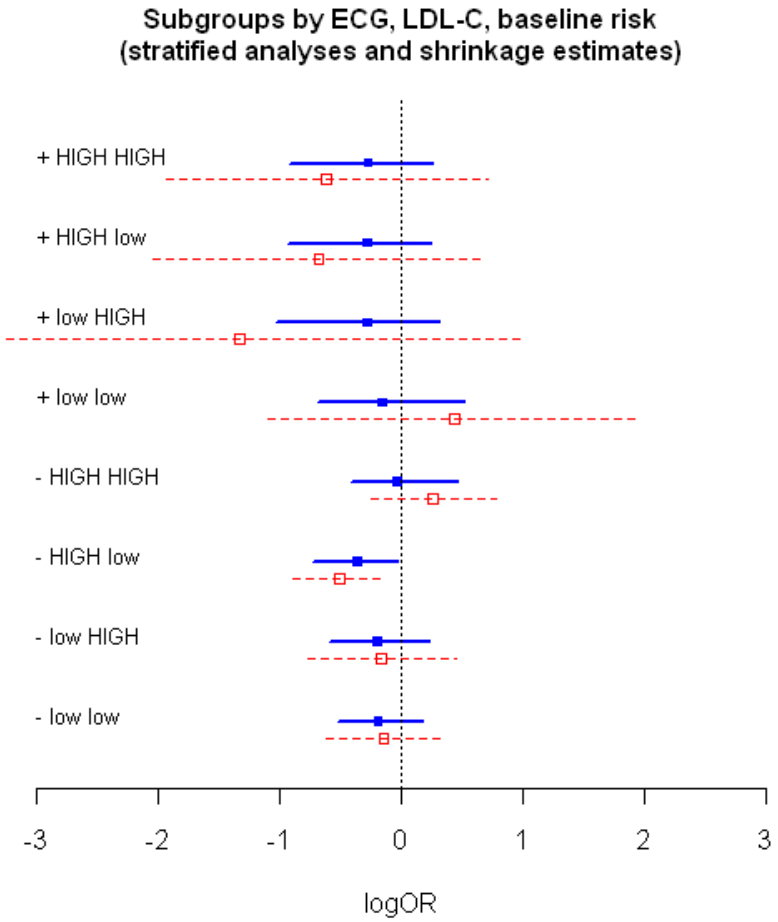
From Davis & Leffingwell (Contr Clinical Trials, 1990)

Note: in the paper a relative risk (using logrank statistic) was used instead of the odds-ratio!

Simple Shrinkage

Example 1 (Davis & Leffingwell 1990): simple shrinkage estimates





Alternative subgroup models

And extensions to meta-analysis

A recap of the subgroup models introduced so far

1) Full stratification $\theta_1, \ldots\ldots, \theta_G$

$\Rightarrow$Assumes a different treatment effect in each subgroup

2) Equal Parameters $\theta_1 = \ldots = \theta_G$

$\Rightarrow$ Assumes the same treatment effect in each subgroup

3) Simple shrinkage estimation $\theta_1, \theta_2, \ldots, \theta_G \sim N(\mu, \omega^2)$

$\Rightarrow$ Assumes exchangeability among the subgroup effects

Issues with simple shrinkage assumption

- Exchangeability for subgroup effects may be questionable
  - In particular if subgroups are defined by covariates that are thought to be predictive of the effects
- Therefore, in this section we look at some alternative approaches to shrinkage that might address this problem in certain circumstances
- Based on the subsequent case-study we will look at the case of 3 binary covariates *A,B,C*, defining *8* subgroups

General first order interaction model with 3 binary covariates

- Effect for subgroup *g*

$$\theta_g = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high)$$

  - τ fixed baseline (all covariates = 0)
  - γ first-order interactions

- If γ's are separate fixed effects we would have a completely standard simple regression model with first order interactions

Simple regression and simple shrinkage

- It is possible to combine simple regression with a simple shrinkage model
- However, the interpretation is a bit strange
- The subgroup effects are exchangeable after accounting for a first order interaction

$$\theta_g = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high) + \varphi_g$$

$\varphi_3 \sim$ Normal$(0, \omega^2)$ with prior on ω

The Dixon-Simon Model

shrinkage on the regression model parameters

- Here we start with the simple regression model

$$\theta_g = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high)$$

  - τ fixed baseline treatment effect

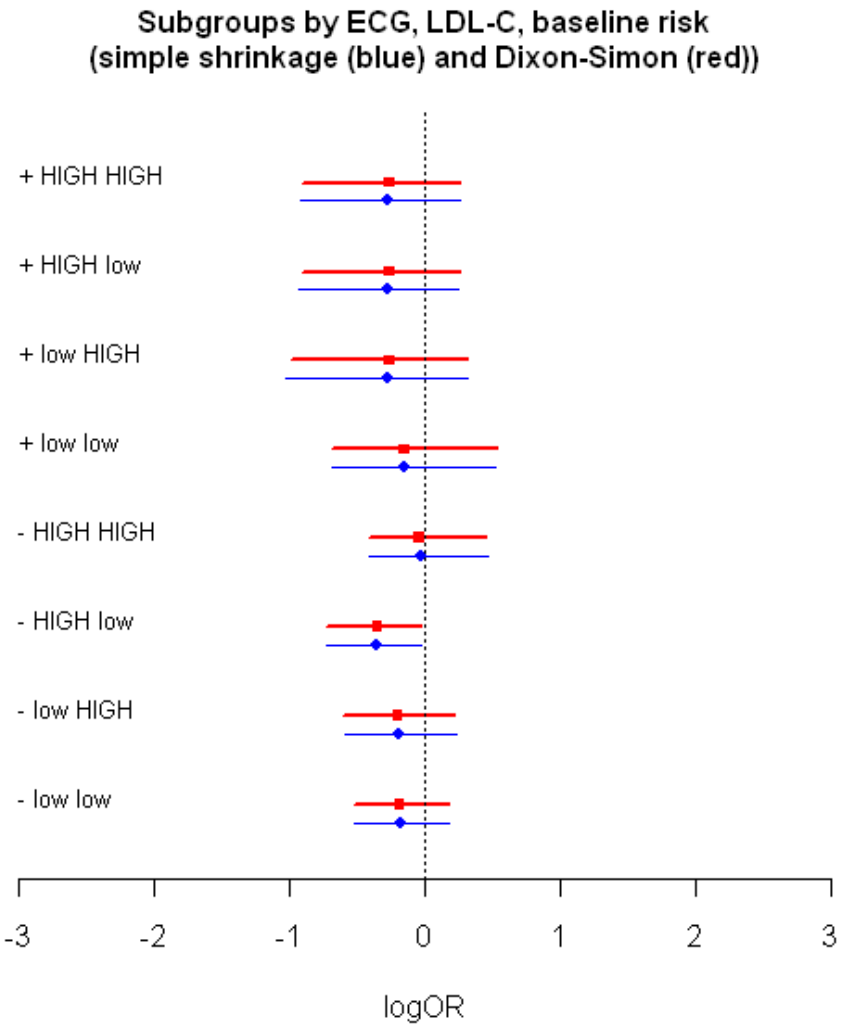
- Shrinkage is then applied to the regression model coefficients:

$$\gamma_1,\ \gamma_2,\ \gamma_3 \sim \text{Normal}(0, \omega^2) \text{ with prior on } \omega$$

- This is similar to penalized regression techniques

Example 1

Simple shrinkage and Dixon-Simon model



Higher order interaction model for 3 binary covariates

- Effect for subgroup *g*

$$\theta_g = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high)$$
$$+ \delta_1 I(A = B = high) + \delta_2 I(A = C = high) + \delta_3 I(B = C = high)$$
$$+ \alpha I(A = B = C = high)$$

  - τ fixed baseline (all covariates = 0)
  - γ first-order interactions
  - δ second-order interaction
  - α third-order interaction

- Note: the full model without any structure on parameters corresponds to a fully stratified analysis (just a reparameterization!)

Extended Dixon and Simon model with higher order interactions

- Effect for subgroup *g*

$$\theta_g = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high)$$
$$+ \delta_1 I(A = B = high) + \delta_2 I(A = C = high) + \delta_3 I(B = C = high)$$
$$+ \alpha I(A = B = C = high)$$

  - τ fixed baseline

- $\gamma_1, \gamma_2, \gamma_3 \sim Normal(0, \omega_1^2)$
- $\delta_1, \delta_2, \delta_3 \sim Normal(0, \omega_2^2)$
- $\alpha \sim$ Normal$(0, \omega_3^2)$
- with priors on $\omega_1, \omega_2, \omega_3$

Meta-analysis: extensions to multiple studies

- Effect for subgroup *g* in study *s*

$$\theta_{gs} = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high) + \lambda_s$$

- Equal Parameters $\lambda_1 = \ldots = \lambda_S$
  - Fixed or common effect meta-analysis assumption
- Exchangeability estimation $\lambda_s \sim \text{Normal}(0, \varphi^2)$, $s=1,\ldots,S$
  - *Random effects meta-analysis assumption*
- Applicable with all subgroup models

Recap on subgroup models

1. Identical subgroup effects
2. Fully stratified analysis
3. Regression structure with first order subgroup interactions, no random effects (regression model)
4. Simple shrinkage (full exchangeability)
5. Regression structure + additonal random effects (partial exchangeability model)
6. Dixon-Simon (first order interactions with shrinkage placed on the coefficients)
7. Extended Dixon-Simon (shrinkage placed on coefficients associated with first and higher order interactions)

Full set of models

(1) No subgroup effect	$\theta_g = \tau, \quad g = 1, \ldots, 8$ $\tau \sim \text{Normal}\,(0, 10^6)$
(2) Fully stratified	$\theta_g = \tau + \gamma_1 I_{\{B_1+\}} + \gamma_2 I_{\{B_2+\}} + \gamma_3 I_{\{B_3+\}}$ $+\delta_1 I_{\{B_1+,B_2+\}} + \delta_2 I_{\{B_1+,B_3+\}}$ $+\delta_3 I_{\{B_2+,B_3+\}} + \alpha I_{\{B_1+,B_2+,B_3+\}}$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\gamma_k \sim \text{Normal}\,(0, 10^6), \quad k = 1, 2, 3$ $\delta_k \sim \text{Normal}\,(0, 10^6), \quad k = 1, 2, 3$ $\alpha \sim \text{Normal}\,(0, 10^6)$
(3) Simple regression	$\theta_g = \tau + \gamma_1 I_{\{B_1+\}} + \gamma_2 I_{\{B_2+\}} + \gamma_3 I_{\{B_3+\}}$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\gamma_k \sim \text{Normal}\,(0, 10^6), \quad k = 1, 2, 3$
(4) Simple shrinkage	$\theta_g = \tau + \psi_g, \quad g = 1, \ldots, 8$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\psi_g \sim \text{Normal}\,(0, \omega^2)$ $\omega \sim \text{Half-normal}\,(1)$
(5) Simple regression and shrinkage	$\theta_g = \tau + \gamma_1 I_{\{B_1+\}} + \gamma_2 I_{\{B_2+\}} + \gamma_3 I_{\{B_3+\}} + \psi_g$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\gamma_k \sim \text{Normal}\,(0, 10^6), \quad k = 1, 2, 3$ $\psi_g \sim \text{Normal}\,(0, \omega^2), \quad g = 1, \ldots, 8$ $\omega \sim \text{Half-normal}\,(1)$
(6) Dixon and Simon	$\theta_g = \tau + \gamma_1 I_{\{B_1+\}} + \gamma_2 I_{\{B_2+\}} + \gamma_3 I_{\{B_3+\}}$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\gamma_k \sim \text{Normal}\,(0, \omega^2), \quad k = 1, 2, 3$ $\omega \sim \text{Half-normal}\,(1)$
(7) Extended Dixon and Simon	$\theta_g = \tau + \gamma_1 I_{\{B_1+\}} + \gamma_2 I_{\{B_2+\}} + \gamma_3 I_{\{B_3+\}}$ $+\delta_1 I_{\{B_1+,B_2+\}} + \delta_2 I_{\{B_2+,B_3+\}}$ $+\delta_3 I_{\{B_1+,B_3+\}} + \alpha I_{\{B_1+,B_2+,B_3+\}}$ $\tau \sim \text{Normal}\,(0, 10^6)$ $\gamma_k \sim \text{Normal}\,(0, \omega_1^2), \quad k = 1, 2, 3$ $\delta_k \sim \text{Normal}\,(0, \omega_2^2), \quad k = 1, 2, 3$ $\alpha \sim \text{Normal}\,(0, \omega_3^2)$ $\omega_l \sim \text{Half-normal}\,(1), \quad l = 1, 2, 3$

Case Study

Case study

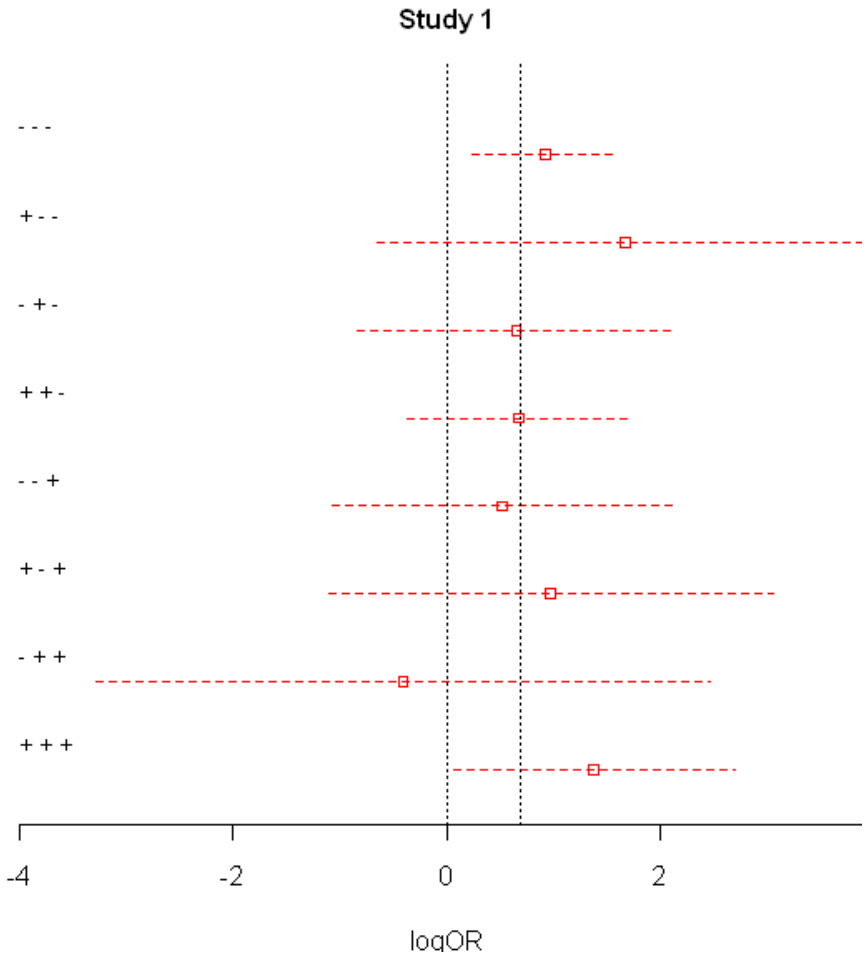
Results

- Separate analyses for two trials
  - “small” trial 1
  - “large” trial 4
- Meta-analytic subgroup analyses: all seven trials
- Results for two models are shown
  - Dixon-Simon: exchangeable 1st order terms
  - extended Dixon-Simon: exchangeable 1st and higher order interaction terms

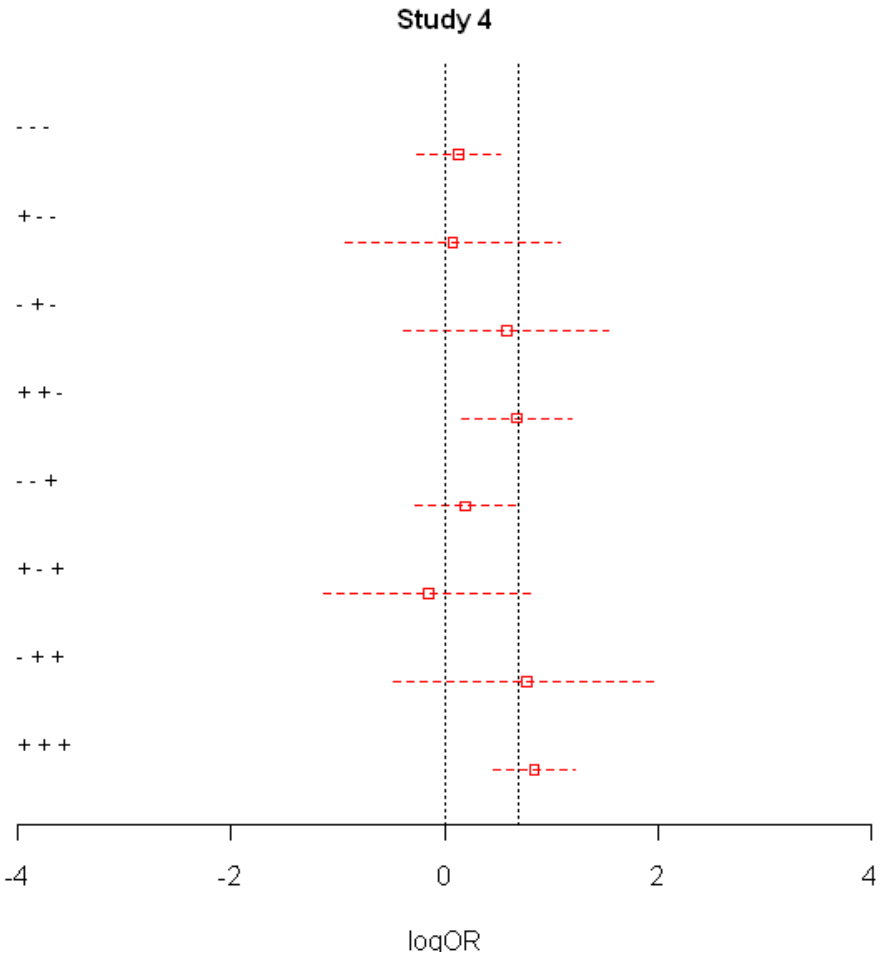
Case Study

… Data for small and large study (study 1 and study 4)

Fully stratified



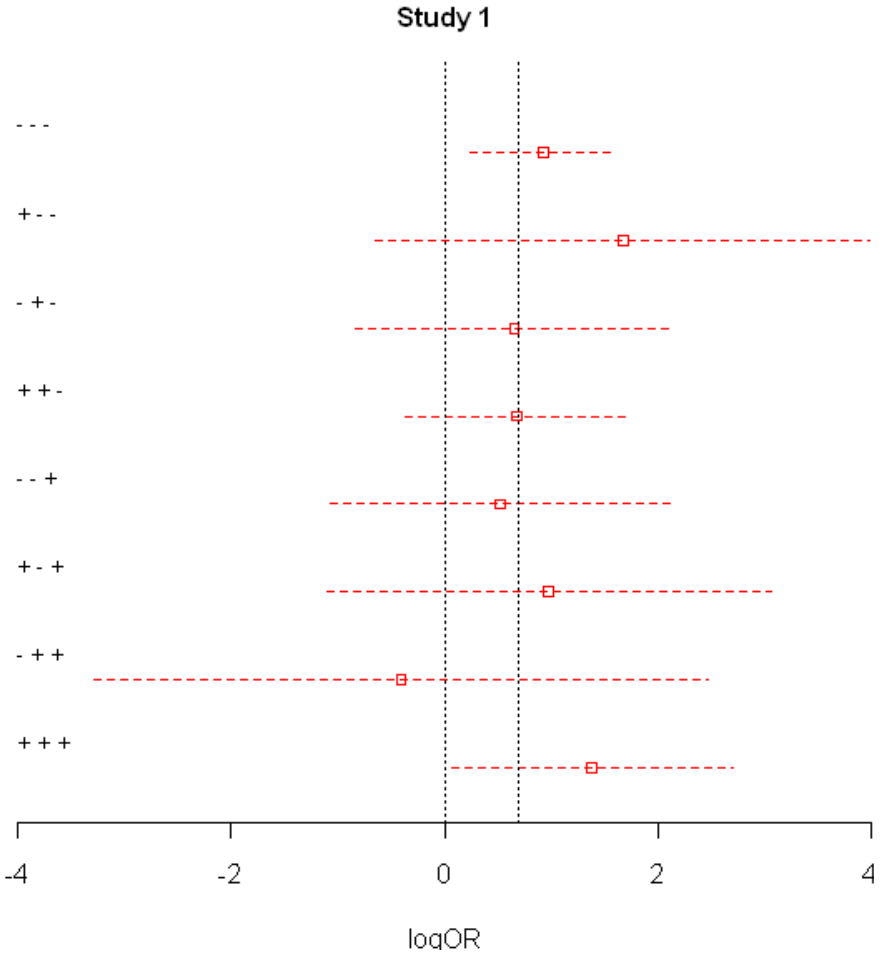
Fully stratified



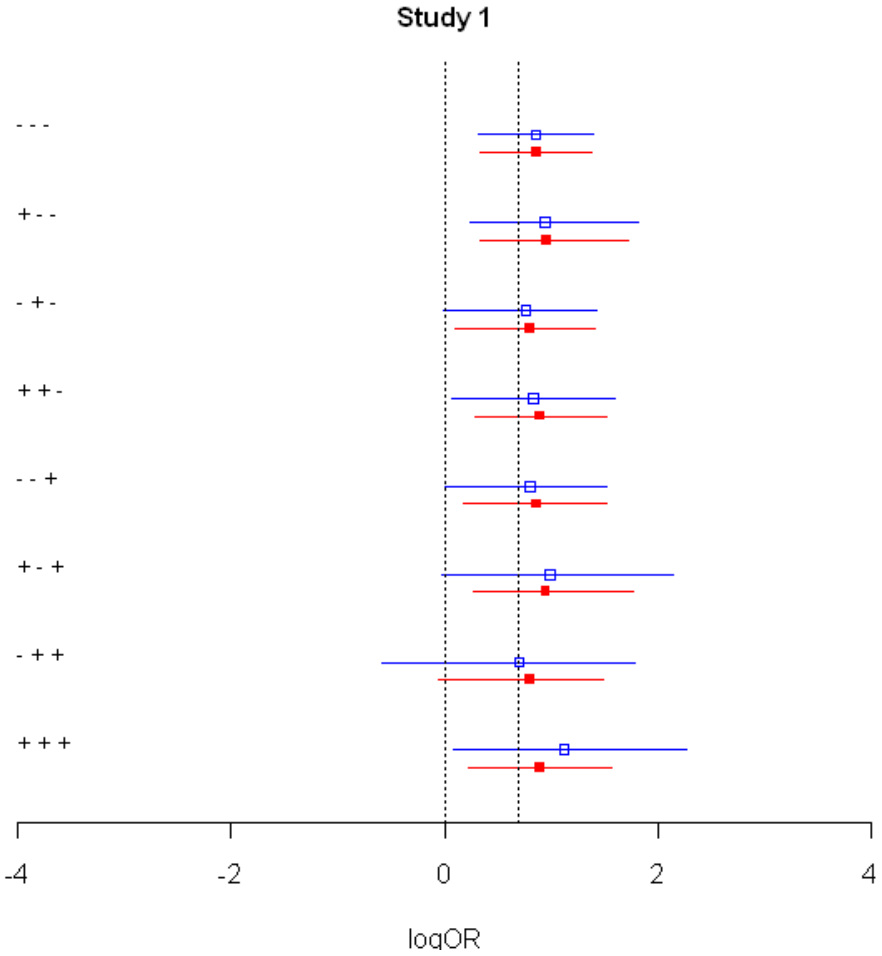
Case Study

Two subgroup analyses for Study 1

Fully stratified



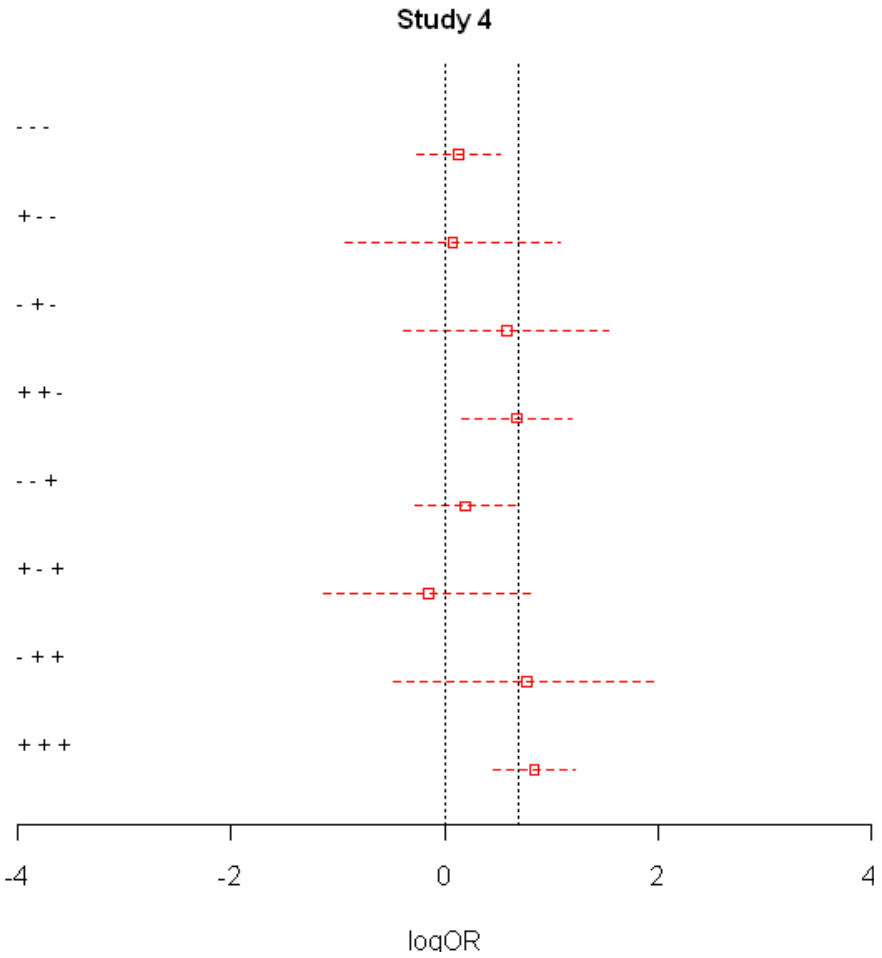
Dixon-Simon
Extended Dixon-Simon



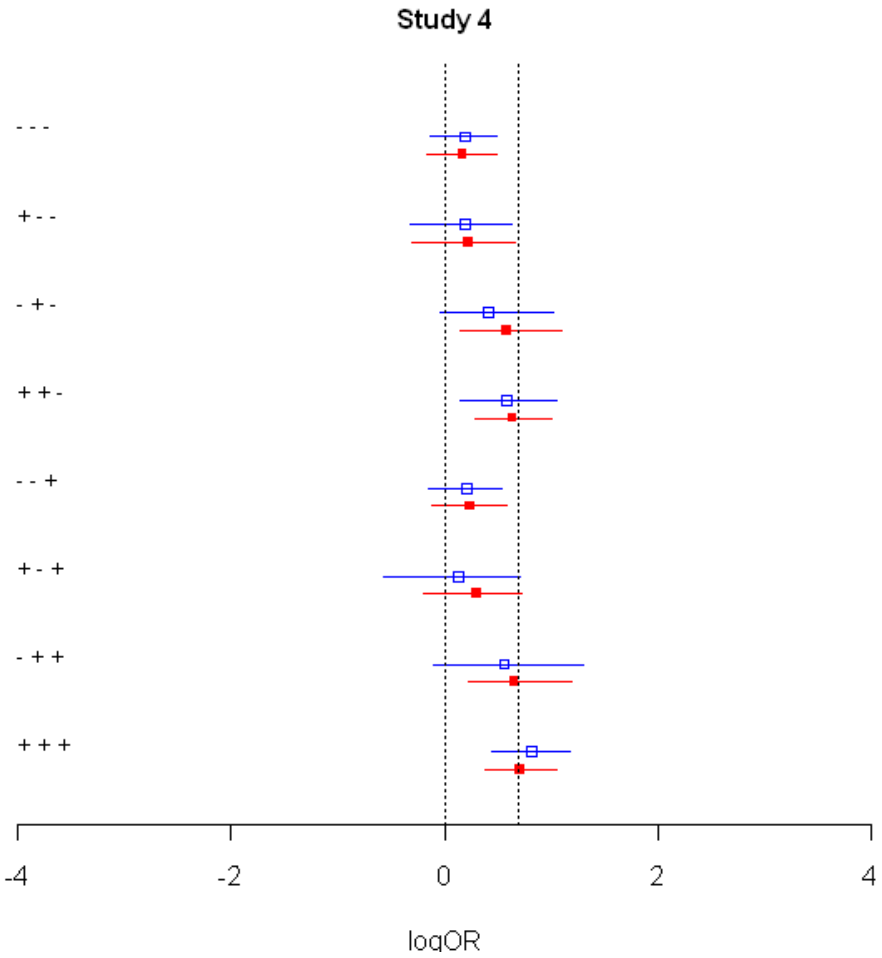
Case Study

Two subgroup analyses for Study 4

Fully stratified

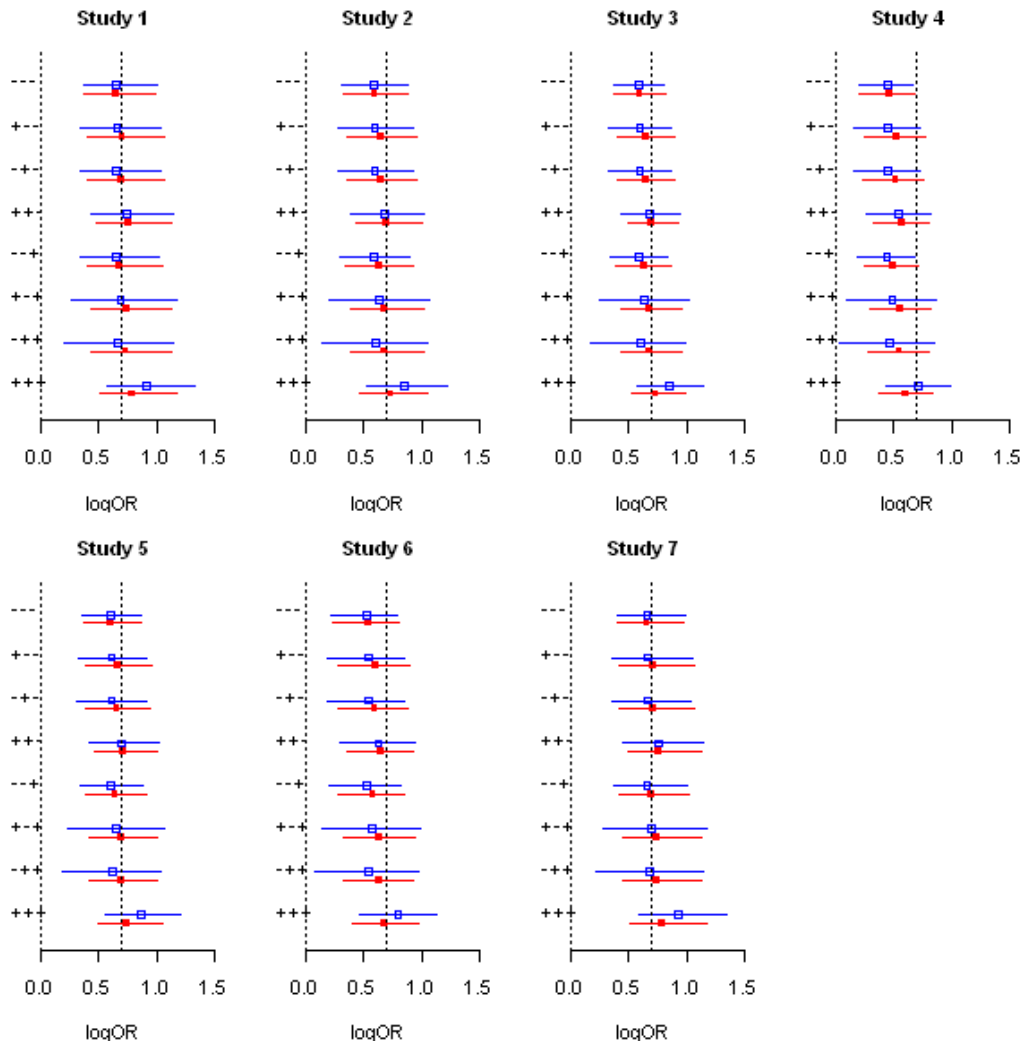


Dixon-Simon
Extended Dixon-Simon



Case Study

Two meta-analytic subgroup analyses



Two models

- Dixon-Simon + study effects (red)
- Extended Dixon-Simon + study effects (blue)
- Both with similar deviance information criterion (DIC)
- Model diagnostics reasonably good
- Qualitatively similar results

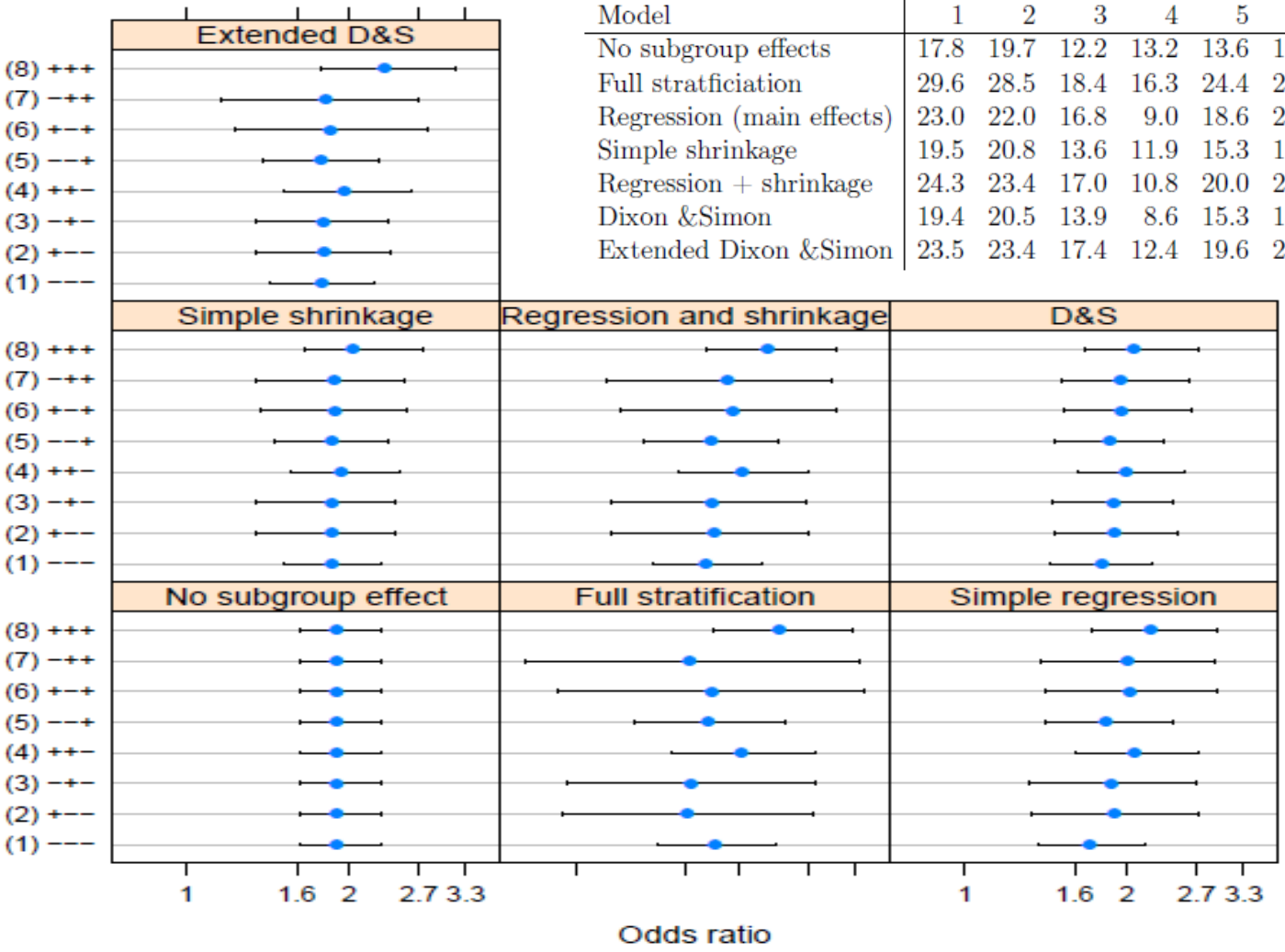
A recap on the modeling strategy

- Priors are carefully selected that we hope are dominated by the data
- A variety of modeling structures examined
  - Model support measured using the deviance information criteria (DIC) Model diagnostics with frequentist properties used to help show whether a model has good calibration
  - Examine if similar conclusions are reached from well supported models to check inference robustness
- This work builds upon the work of Box (1980), who advocated the use of an iterative cycle of model criticism and estimation

Sensitivity analyses across a range of structures

Using DIC for model comparison

Model	Trial 1	2	3	4	5	6	7	Meta-Analysis
No subgroup effects	17.8	19.7	12.2	13.2	13.6	17.0	19.9	92.3
Full stratficiation	29.6	28.5	18.4	16.3	24.4	27.0	31.5	101.3
Regression (main effects)	23.0	22.0	16.8	9.0	18.6	20.2	24.7	95.2
Simple shrinkage	19.5	20.8	13.6	11.9	15.3	18.6	21.5	93.7
Regression + shrinkage	24.3	23.4	17.0	10.8	20.0	21.6	25.8	96.9
Dixon &Simon	19.4	20.5	13.9	8.6	15.3	18.3	21.3	93.1
Extended Dixon &Simon	23.5	23.4	17.4	12.4	19.6	22.7	25.4	96.3



Concluding Remarks

Concluding Remarks

- Post-hoc subgroup analyses with a small number of subgroups defined by clinically important baseline factors
- Testing approaches have clear limitations due to small sample sizes and multiplicity problems
- Inferential/estimation approaches based on shrinkage ideas are more promising
- Required: a “model” for the similarity of subgroup effects
  - Simple shrinkage model
  - Dixon-Simon model or extended version(s)
- Examples: different shrinkage models lead to similar answers

Part 4 safety network meta-analysis

Acknowledgements

Based on the work of the Bayesian DIA safety meta-analysis team

Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis

David Ohlssen[1]*, Karen L. Price[2], H. Amy Xia[3], Hwanhee Hong[4], Jouni Kerman[5], Haoda Fu[2], George Quartey[6], Cory R. Heilmann[2], Haijun Ma[3], and Bradley P. Carlin[4]

[1]Novartis Pharmaceuticals Corporation

[2]Eli Lilly and Company

[3]Amgen, Inc

[4]Division of Biostatistics, University of Minnesota

[5]Novartis Pharma AG

[6]Genentech, Inc

Research Report 2013-5, Division of Biostatistics, University of Minnesota, 2013, Submitted to *Pharmaceutical Statistics*.

Introduction to Bayesian Network Meta-Analysis

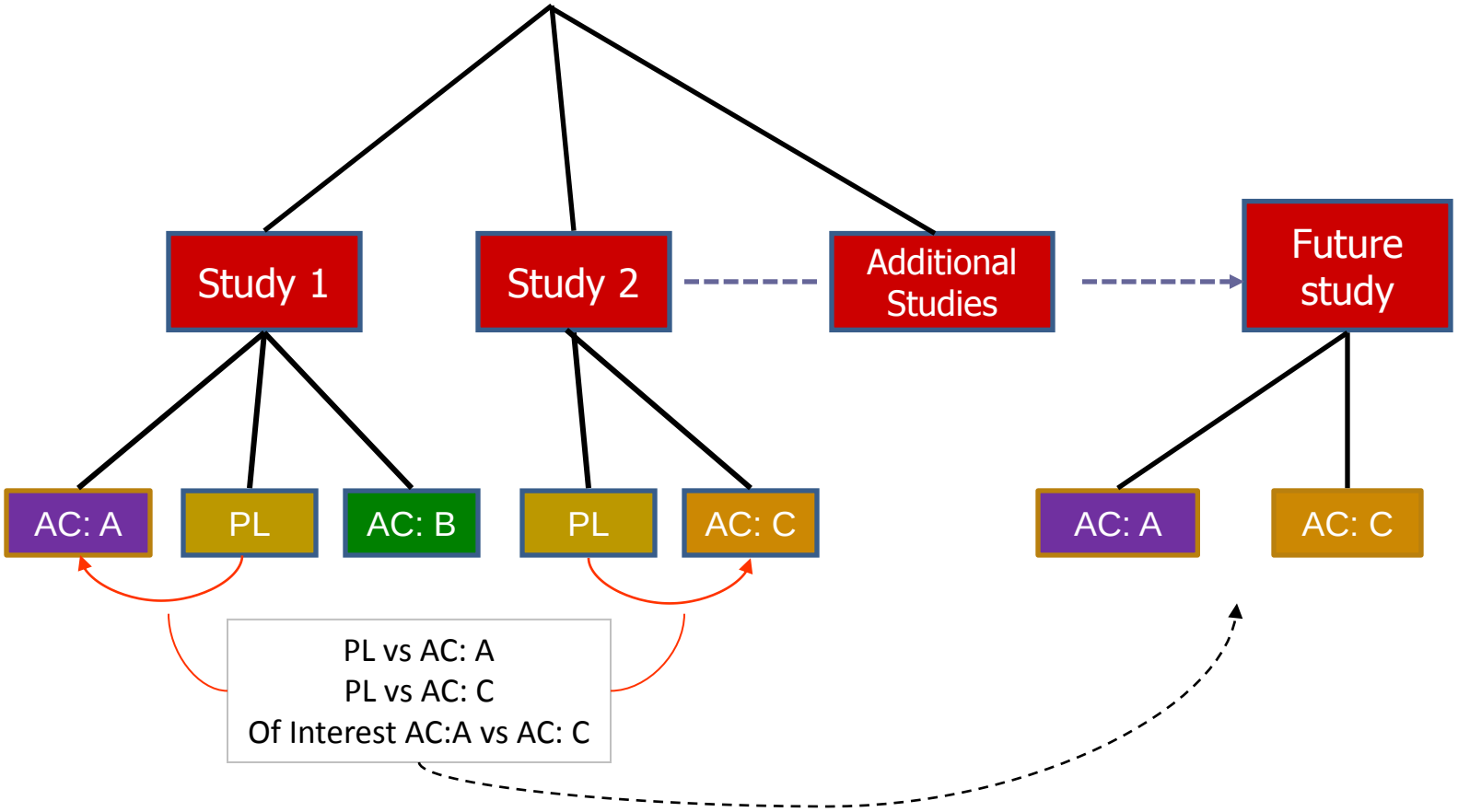
Bayesian Network Meta-Analysis

- Systematic reviews are considered standard practice to inform evidence-based decision-making regarding efficacy and safety
- Bayesian network meta-analysis (mixed treatment comparisons) have been presented as an extension of traditional MA by including multiple different pairwise comparisons across a range of different interventions
- Several Guidances/Technical Documents recently published

Example References

- ISPOR: Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health Care Decision-making
- ISPOR: Conducting Indirect Treatment Comparisons and Network Meta-Analysis for Health Care Decision-making
- NICE DSU Technical Support Documents
- Canadian Agency for Drugs and Technologies in Health Report
- Spiegelhalter, Abrams, Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley 2003

Basic Framework



AC: Active Comparator

Poisson network meta-analysis model

Based on the work of Lu and Ades (2006 & 2009)

$$r_{jk} \sim \text{Poisson}(\lambda_{jk} E_{jk}) \quad j = 1,\ldots,M;\ k = 1,\ldots,K$$

$$\log(\lambda_{jb}) = \mu_j$$

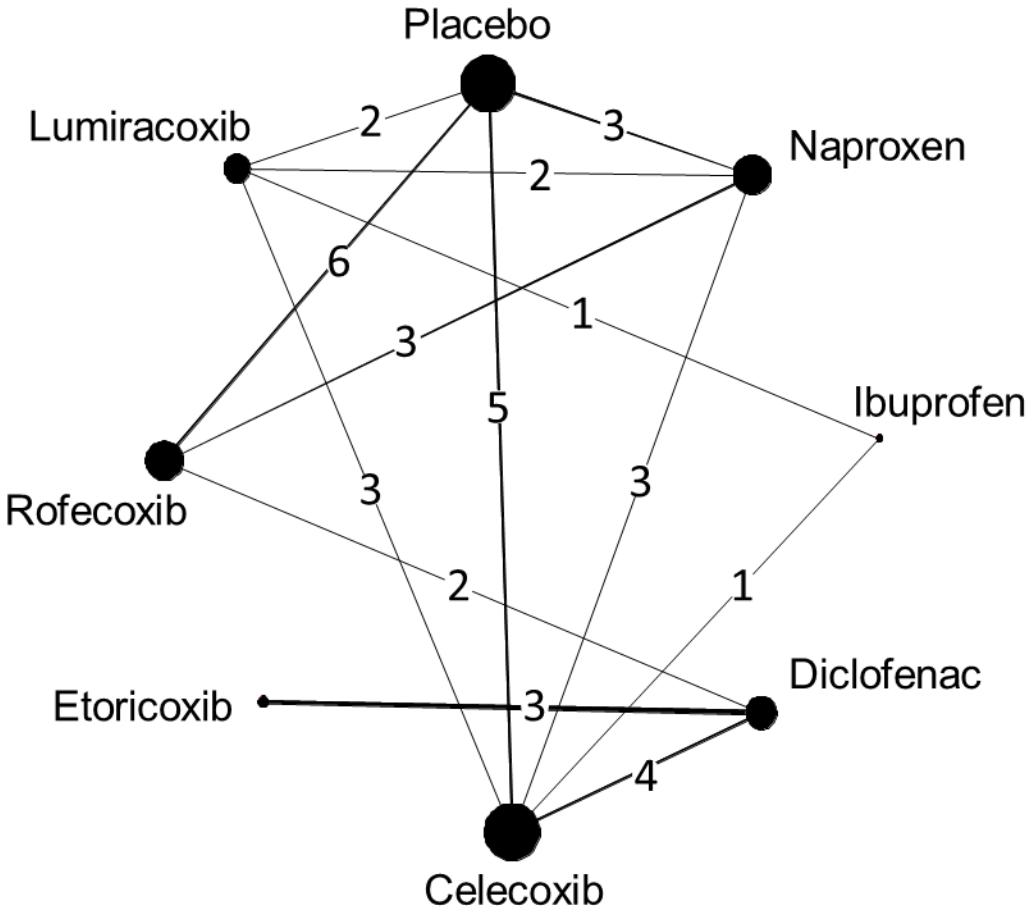
$$\log(\lambda_{jk}) = \mu_j + \delta_{jbk} \quad k \neq b$$

b is the control treatment associated with trial j

- μ_j is the effect of the baseline treatment *b in trial j* and δ_{jbk} is the trial-specific treatment effect of treatment *k* relative to treatment to b (the baseline treatment associated with trial j)
- Note baseline treatments can vary from trial to trial
- Different choices for μ's and δ's. They can be: *common* (over studies), *fixed* (unconstrained), or "*random*"
- Consistency assumptions required among the treatment effects
- Prior distributions required to complete the model specification

Network meta-analysis

Trelle et al (2011) - Cardiovascular safety of non-steroidal anti-inflammatory drugs:



- Primary Endpoint was myocardial infarction
- Data synthesis 31 trials in 116 429 patients with more than 115 000 patient years of follow-up were included.
- A Network random effects meta-analysis were used in the analysis
- Critical aspect – the assumptions regarding the consistency of evidence across the network
- How reasonable is it to rank and compare treatments with this technique?

Trelle, Reichenbach, Wandel, Hildebrand, Tschannen, Villiger, Egger, and Juni. Cardiovascular safety of non-steroidal anti-inflammatory drugs network meta-analysis. *BMJ* 2011; 342: c7086. Doi: 10.1136/bmj.c7086

Results from Trelle *et al*
Myocardial infarction analysis

Relative risk with 95% confidence interval compared to placebo

Treatment	RR estimate	lower limit	upper limit
Celecoxib	1.35	0.71	2.72
Diclofenac	0.82	0.29	2.20
Etoricoxib	0.75	0.23	2.39
Ibuprofen	1.61	0.50	5.77
Lumiracoxib	2.00	0.71	6.21
Naproxen	0.82	0.37	1.67
Rofecoxib	2.12	1.26	3.56

Authors' conclusion:

Although uncertainty remains, little evidence exists to suggest that any of the investigated drugs are safe in cardiovascular terms. Naproxen seemed least harmful.

Comments on Trelle *et al*

- Drug doses could not be considered (data not available).
- Average duration of exposure was different for different trials.
- Therefore, ranking of treatments relies on the strong assumption that the risk ratio is constant across time for all treatments
- The authors conducted extensive sensitivity analysis and the results appeared to be robust

Two way layout via MAR assumption

- An alternative way to parameterize proposed by Jones et al (2011) and Piephoetal *et al* (2012) uses a classical two-way linear predictor with main effects for treatment and trial.
- Both papers focus on using the two-way model in the classical framework. By using the MAR property a general approach to implementation in the Bayesian framework can be formed
- All studies can in principle contain every arm, but in practice many arms will be missing. As the network meta-analysis model implicitly assume MAR (Lu and Ades; 2009) a common (though possibly missing) baseline treatment can be assumed for every study (Hong and Carlin; 2012)

$$\text{logit}\left(p_{ik}\right) = s_i + t_k + v_{ik}$$

$$\begin{aligned} t_1 &= 0 \\ v_{i1} &= 0 \; i = 1, \ldots, N \end{aligned}$$

$$\left(v_{i2}, \ldots, v_{iK}\right)' \sim \text{MVN}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

Comments on implementation and practical advantages

- In WinBUGS include every treatment in every trial with missing outcome cells for missing treatments
- Utilize a set of conditional univariate normal distributions to form the multivariate normal (this speeds up convergence)
- The parameterization has several advantages when forming priors:
  - In the Lu and Ades model, default “non-informative” priors must be used as the trial baseline parameters are nuisance parameters with no interpretation
  - In the two-way model an informative prior for a single treatment baseline treatment can be formed as each trial has the same parameterization
  - In the two way model there is much greater control over non-informative priors. This can be valuable when you have rare safety events asymmetry in prior information can potentially lead to a bias

Alternative approach Full multivariate meta-analysis

- Instead of associating a concurrent control parameter with each study, an alternative approach is to place random effects on every treatment main effect
- This creates a so called multivariate meta-analysis

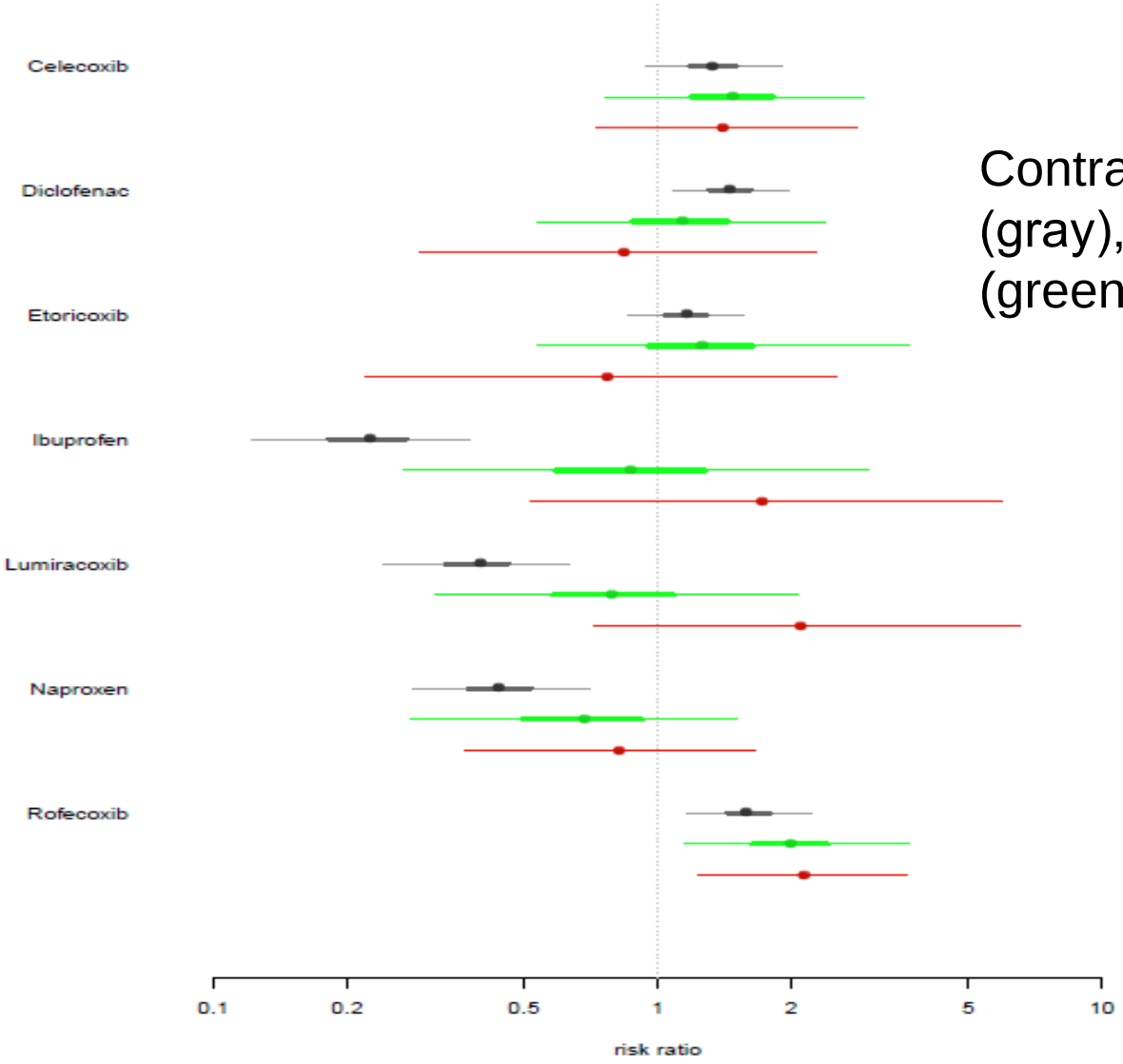
$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(n_{jk}, p_{jk}\right) \quad j = 1, \ldots, m \;\; k = 1, \ldots, K \\ \text{Logit}\left(p_{jk}\right) &= \gamma_{jk} \end{aligned}$$

$$\begin{pmatrix} \gamma_{j1} \\ \vdots \\ \gamma_{jK} \end{pmatrix} \sim \text{MVN}\left(\left(\mu_1, \ldots, \mu_k\right)', \boldsymbol{\Sigma}\right).$$

$$p\left(\boldsymbol{\mu}\right), \;\; p\left(\boldsymbol{\Sigma}\right)$$

MI results from Trelle et al

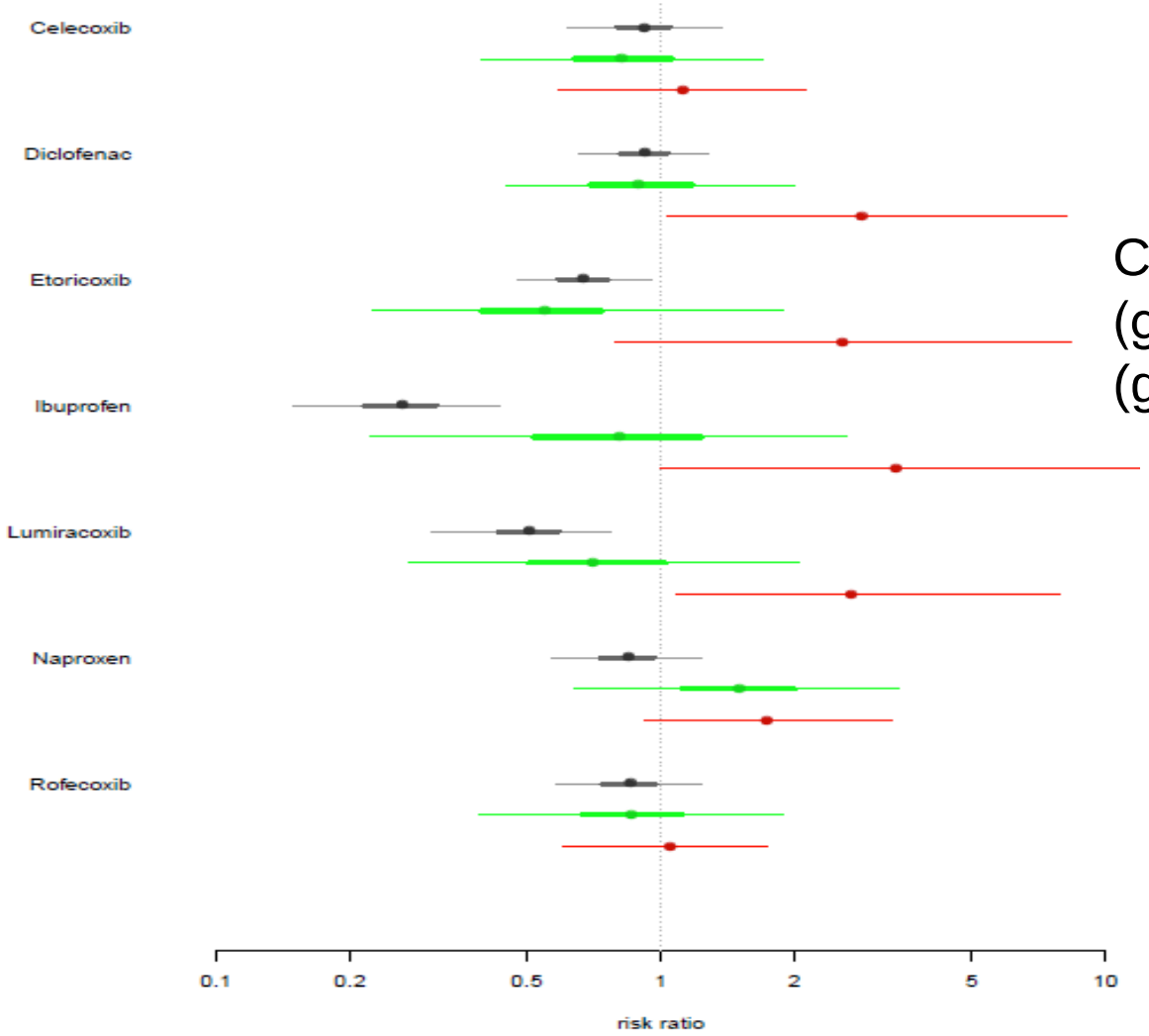
Comparing Bristol RE model with multivariate random effects



Contrasts to placebo: Pooled (gray), Arm−based MV model (green), Trelle (red)

Stroke results from Trelle et al

Comparing Bristol RE model with multivariate random effects



Contrasts to placebo: Pooled (gray), Arm−based MV model (green), Trelle (red)

Discussion of full multivariate meta-analysis model

- Allows borrowing of strength across baseline as every treatment is considered random
- Therefore, in rare event meta-analysis, incorporates trials with zero total events through the random effects
- No consistency relations to deal with!
- Priors on the variance components can be formed using inverse Wishart or using Cholesky decomposition
- Breaks the concurrent control structure so automatically will introduce some confounding

New challenges

- Network meta-analysis with multiple outcomes
  - Sampling model (multinomial?)
  - Borrow strength across treatment effects
  - Surrogate outcome meta-analysis combined with a network meta-analysis
- Network meta-analysis with subgroup analysis
- Combining network meta-analysis; meta-analysis of subgroups and multivariate meta-analysis

Overall conclusions

- Many opportunities for Bayesian methods to help handle selection problems in drug development
- Bayesian approaches to hypothesis testing appear to provide an attractive way to detect signals
- However, in practice models with strong structural assumptions and or informative priors are often required
- Therefore, I prefer estimation based techniques that help characterize heterogeneity and help assess reproducibility
- These techniques:
  - Should be backed up with model sensitivity analysis
  - Require going well beyond statistics to make final decisions

References motivating examples

ICH E5 Ethnic Factors in the Acceptability of Foreign Clinical Data. European Medicines Agency.September 1998. Availablehttp://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002842.pdf

Food and Drug Administration Cardiovascular and Renal Drugs Advisory Committee Meeting, July 28, 2010. AstraZeneca briefing materials. Available at http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM220197.pdf

Bombardier C, Laine L, Reicin A, et al for the VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000; **343:** 1520–28.

Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med 2005;352:1092-1102

Kevin J. Carroll and Thomas R. Fleming. Statistical evaluation and analysis of regional interactions: he PLATO trial case study. (in press) Statistics in Biopharmaceutical Research

Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004;364(9450):2021-9.

References – overview of Bayesian techniques

- Bayarri, M. and Morales, J. (2003) Bayesian measures of surprise for outlier detection. *J. Statist. Planng Inf.*, **111**, 3-22
- Bayarri, M. J. and Castellanos, M. E. (2007). Bayesian Checking of the Second Levels of Hierarchical Models. Statistical Science 22 322–343.
- Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. Biometrics Jun 2004; 60(2):418–426, doi:10.1111/j.0006-341X.2004.00186.x. PMID:15180667.
- Dumouchel W. Multivariate Bayesian logistic regression for analysis of clinical study safety issues. Statistical Science 2012; 27(3):319–339, doi:10.1214/11-STS381.
- Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990; **9:** 657–71.
- Jones, H. E., Ohlssen, D. I. and Spiegelhalter, D. J. (2008). Use of the false discovery rate when comparing multiple health care providers. Journal of Clinical Epidemiology 61 232–240.e2.
- Ohlssen, D. I., Sharples, L. D. and Spiegelhalter, D. J. (2007). A hierarchical modelling framework for identifying unusual performance in health care providers. Journal of the Royal Statistical Society: Series A (Statistics in Society) 170 865–890.
- Marshall, E. C. and Spiegelhalter, D. J. (2007). Identifying outliers in Bayesian hierar- chical models: a simulation-based approach. Bayesian Analysis 2 409–444.
- Sivaganesan S, Laud PW, and Müller P. A Bayesian subgroup analysis with a zero-enriched Polya Urn scheme. Statistics in Medicine 2011, 30:312—323

Subgroup analyses references

Davis, Leffingwell (1990). Empirical Bayes estimates of subgroup effects in clinical trial. *Controlled Clinical Trials*, 11: 37-42

DerSimonian, Laird (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7; 177-88

Dixon, Simon (1991). Bayesian subgroup analysis. *Biometrics*, 47: 871-81

Fleming (2010), "Clinical Trials: Discerning Hype From Substance," *Annals of Internal Medicine* 153:400 -406.

Hodges, Cui, Sargent, Carlin (2007). Smoothing balanced single-error terms Analysis of Variance. *Technometrics*, 49: 12-25

Jones, Ohlssen, Neuenschwander, Racine, Branson (2011). Bayesian models for subgroup analysis in clinical trials. *Clinical Trials Clinical Trials* 8 129 -143

Louis (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *JASA*, 79: 393-98

Pocock, Assman, Enos, Kasten (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practic eand problems. Statistics in Medicine, 21: 2917–2930

- Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley: Chichester, 2004.

Thall, Wathen, Bekele, Champlin, Baker, Benjamin (2003). Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes, *Statistics in Medicine*, 22: 763-80

Network meta-analysis references

- Dias S, Welton N, Sutton A, Ades A. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials 2011. Available at Available at: http://www.nicedsu.org.uk (accessed 09.27.2012).
- Hoaglin D, Hawkins N, Jansen J, Scott D, Itzler R, Cappelleri J, Boersma C, Thompson D, Larholt K, Diaz M, et al.. Conducting Indirect-Treatment-Comparison and Network-Meta-Analysis Studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. Value in Health 2011; 14(4):429–437.
- Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine 2004; 23(20):3105–3124.
- Lu G, Ades A. Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association 2006; 101(474):447–459.
- Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009; 10(4):792–805.
- Trelle S, Reichenbach S, Wandel S, Hildebrand P, Tschannen B, Villiger P, Egger M, Jüni P, et al.. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ (Clinical esearch ed.) 2011; 342:c7086.