Subgroup analysis using Bayesian hierarchical models: a case study

Beat Neuenschwander
Amy Racine, David Ohlssen, Michael Branson
and
Hayley Jones (MRC, Cambridge)
Basel, 26 June 2009

NOVARTIS

Outline

- Introduction to subgroup analysis
- Shrinkage
- Models
- Case Study
- Concluding Remarks

NOVARTIS

Introduction to Subgroup Analysis

NOVARTIS

Introduction

- For biological reasons treatments may be more effective in some populations of patients
- Important baseline factors
  - Risk factors
  - Genetic factors
  - Demographic factors

NOVARTIS

Introduction
Various Aspects

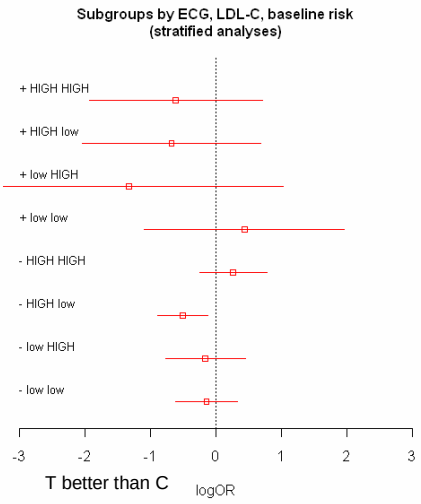
(Focus of this talk in **bold**)

- 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**
- **...**

NOVARTIS

Example 1
Data from one study



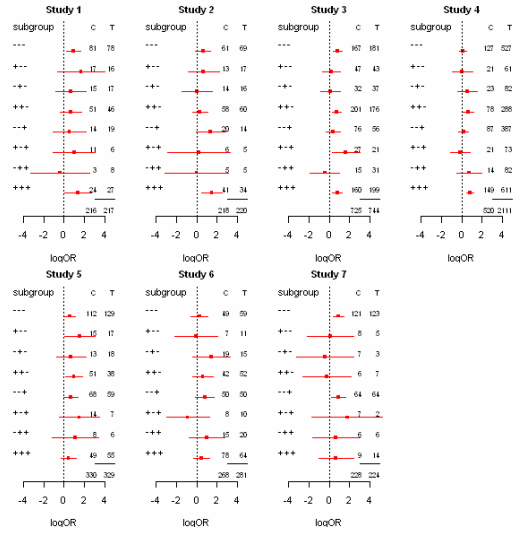
(Davis & Leffingwell, Contr Clin Trials 1990)

- 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)

NOVARTIS

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

NOVARTIS

Approaches

Testing / Estimation

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

NOVARTIS

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 multiplicity problem

NOVARTIS

Testing approaches

Issues

- Post-hoc analyses suffer from
  - small sample sizes due to splitting up the data into subgroups
  - multiplicity problem
- This leads to
  - low power and
  - even wider confidence intervals (due to multiplicity adjustments) compared to fully stratified analysis

NOVARTIS

Estimation Approaches

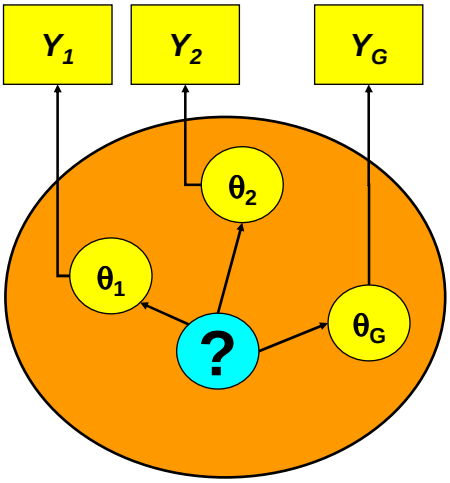
- Various approaches to estimate subgroup effects
- Instead of looking at subgroups in a fully stratified way, it is assumed that information from other subgroups carries information about subgroup(s) of interest
- Subgroup effects $\theta_1, \theta_2, \ldots, \theta_G$ are related/similar to a certain 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

NOVARTIS

Shrinkage

NOVARTIS

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

NOVARTIS

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
  - $Y_g \sim N(\theta_g, s_g^2)$, $g = 1,\ldots,G$
  - $\theta_1, \theta_2,\ldots, \theta_G \sim N(\mu, \tau^2)$
  - 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)

NOVARTIS

Shrinkage

Random effects: sampling interpretation

- Model for similarity: $\theta_1,\ldots,\ \theta_G \sim F$
  - where F is an unknown distribution, e.g. $N(\mu,\ \tau^2)$
- Sampling interpretation
  - True parameters are a sample from an underlying "population"
  - This is somewhat difficult to justify:

    Are the selected subgroups a sample from a population of subgroups?

NOVARTIS

Shrinkage

Random Effects: exchangeability interpretation

- Model for similarity: $\theta_1,\ldots,\ \theta_G \sim F$
- Exchangeability interpretation
  - assumption that the joint probability distribution of $\theta_1,\ldots,\ \theta_G$ is invariant under permutations of the indices $1,\ldots,G$

    (This requires the willingness to talk about the parameters in a fully probabilistic way!)
  - *de Finetti Theorem*: there is a distribution F such that

    $\theta_1,\ldots,\ \theta_G \sim F(\eta)$, *i.e.,* $\theta_1,\ldots,\ \theta_G$ are iid given $F(\eta)$, and $\eta \sim P$ ("prior")
  - There is no sampling interpretation needed here, but an indifference statement about the underlying parameters. A judgment call!
  - Of course we don't know what F is!
- Note:
  - we constantly use exchangeability assumptions about observations
  - for parameters this is less common (except in Bayesian framework)

NOVARTIS

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

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

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

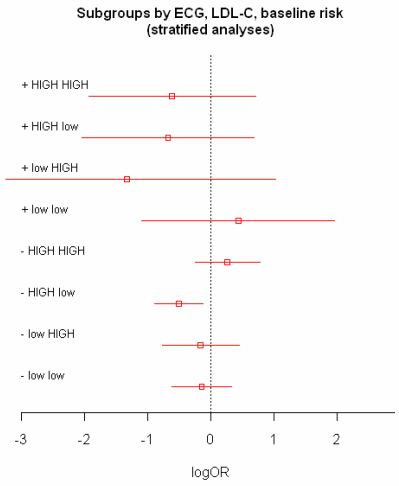
From Davis & Leffingwell (Contr Clinical Trials, 1990)

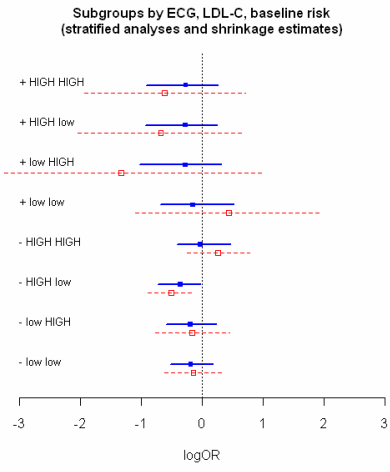
Note: in the paper a relative risk (using logrank statistic) was used instead of the odds-ratio!

NOVARTIS

Simple Shrinkage

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





NOVARTIS

Models

NOVARTIS

Simple Shrinkage Model

Standard Calculus for Normal Data (known τ)

- Data $Y_g \sim N(\theta_g, s_g^2) \quad g = 1,...,G$
- Parameters $\theta_1,...,\theta_G \sim N(\mu, \tau^2)$
- Shrinkage: estimate for each study-specific effect θ_g is
  - a weighted average of study-specific and overall mean estimate

$$\hat{\theta}_g = (1 - B_g)Y_g + B_g\hat{\mu} \qquad \hat{\mu} = \frac{\Sigma_g w_g Y_g}{\Sigma_g w_g} \qquad w_g = (s_g^2 + \tau^2)^{-1}$$

  - with shrinkage factor

$$B_g = \frac{s_g^2}{s_g^2 + \tau^2}$$

    - Small standard error (high precision) $\Rightarrow$ little shrinkage
    - $\tau = 0$ (homogeneity) $\Rightarrow B = 1 \Rightarrow$ complete shrinkage (pooling)
    - τ very large $\Rightarrow B = 0 \Rightarrow$ no shrinkage (complete stratification)

NOVARTIS

Simple Shrinkage Model

Bayesian Formulas for Normal Data (known τ), uniform prior for μ

- overall mean μ is normally distributed with

$$E(\mu \mid Y, \tau) = \frac{\Sigma_g w_g Y_g}{\Sigma_g w_g} \qquad Var(\mu \mid Y, \tau) = \frac{1}{\Sigma_g w_g}$$

- Subgroup effect θ_g is normally distributed with

$$E(\theta_g \mid Y, \tau) = (1 - B_g) Y_g + B_g E(\mu \mid Y, \tau)$$

$$Var(\theta_g \mid Y, \tau) = B_g \left\{ \tau^2 + B_g Var(\mu \mid Y, \tau) \right\}$$

NOVARTIS

Issues

- Even inference for the simplest model is challenging because τ is unknown
  - Classical ways to address this
  - Bayesian approach requires a prior for τ. Inference is automatic/unique, but prior sensitivity should be assessed.
- Exchangeability for subgroup effects may be questionable
  - In particular if subgroups are defined by covariates that are thought to be predictive of the effects
  - We will look at the case of 3 binary covariates *A,B,C*, defining *8* subgroups

NOVARTIS

General 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!)

NOVARTIS

The Dixon-Simon Model

Dixon & Simon, Biometrics, 1990

- 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
  - Dixon-Simon: $\delta_1 = \delta_2 = \delta_3 = \alpha = 0$
  - $\gamma_1, \gamma_2, \gamma_3 \sim$ Normal$(0, \omega^2)$ with prior on ω

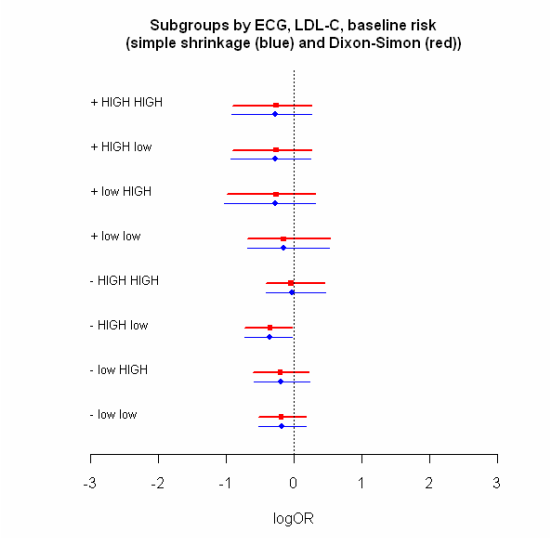
Dixon-Simon Model

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

NOVARTIS

Example 1

Simple shrinkage and Dixon-Simon model



NOVARTIS

Extension of the Dixon-Simon Model

- 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 $\omega_1, \omega_2, \omega_3$
  - Possible constraints: $\omega_1 > \omega_2 > \omega_3$ (lower order interactions typically larger than higher order interactions)

NOVARTIS

Sargent & Hodges Model

… for subgroups defined by 3 binary covariates

- SANOVA (Smooth ANOVA) approach
  - Hodges, Sargent, Cui, Carlin (Technometrics 2007)
- Effect for subgroup *g*

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

  - Each of the 8 regression coefficients assumed independent

$$\text{Normal}(0, \omega_j^2),\ j = 1, \ldots, 8$$

  - But the estimated variance components will be strongly driven by the hyperpriors

NOVARTIS

Extensions to multiple studies

… for subgroups defined by 3 binary covariates

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

$$\theta_{gs} = \tau + \gamma_1 I(A = high) + \gamma_2 I(B = high) + \gamma_3 I(C = high) + \delta_1 I(A = B = high) + \delta_2 I(A = C = high) + \delta_3 I(B = C = high) + \alpha I(A = B = C = high) + \lambda_s$$

  - with exchangeable study effects $\lambda_s \sim \text{Normal}(0, \varphi^2)$, $s = 1, \ldots, S$
  - … and various possible assumptions about the other parameters (Dixon-Simon, extended Dixon-Simon, …)

NOVARTIS

Case Study

NOVARTIS

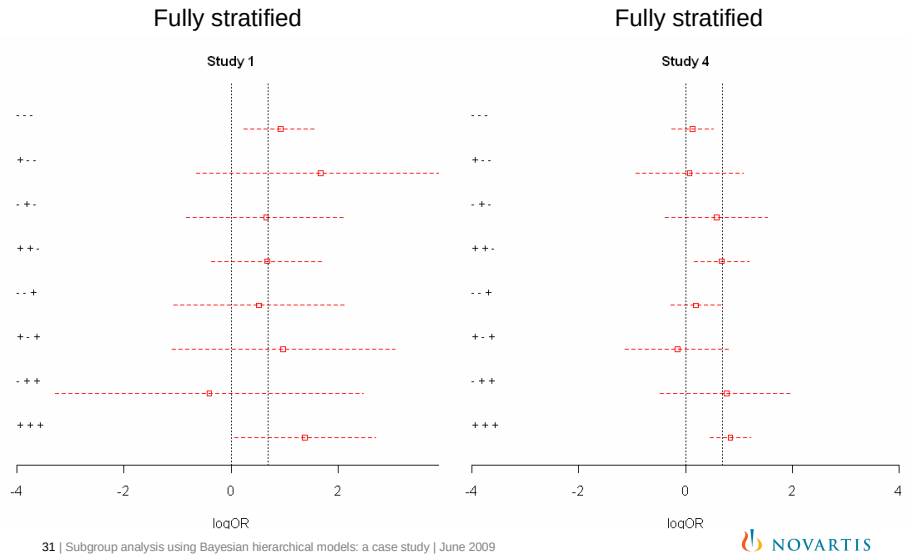
Case study
Results

- 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

NOVARTIS

Case Study

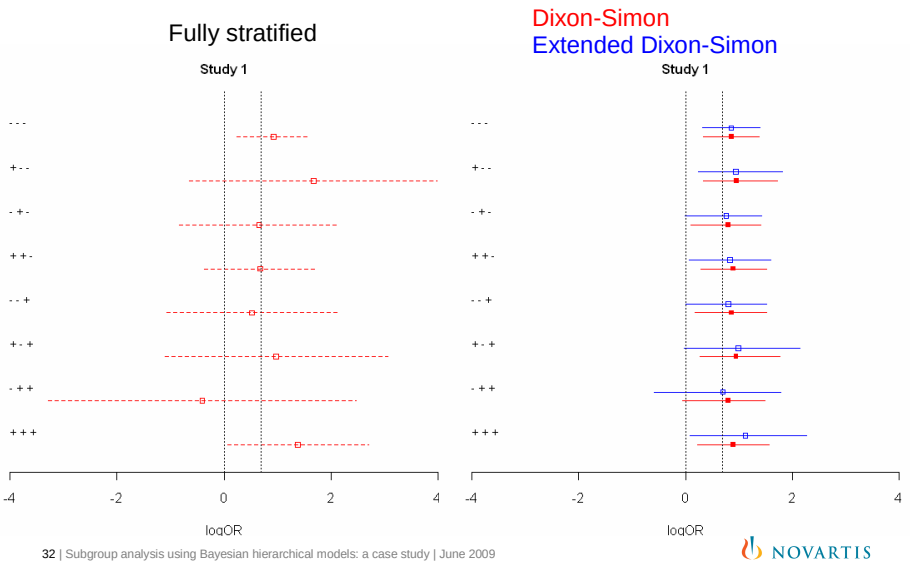
… Data for small and large study (study 1 and study 4)



NOVARTIS

Case Study

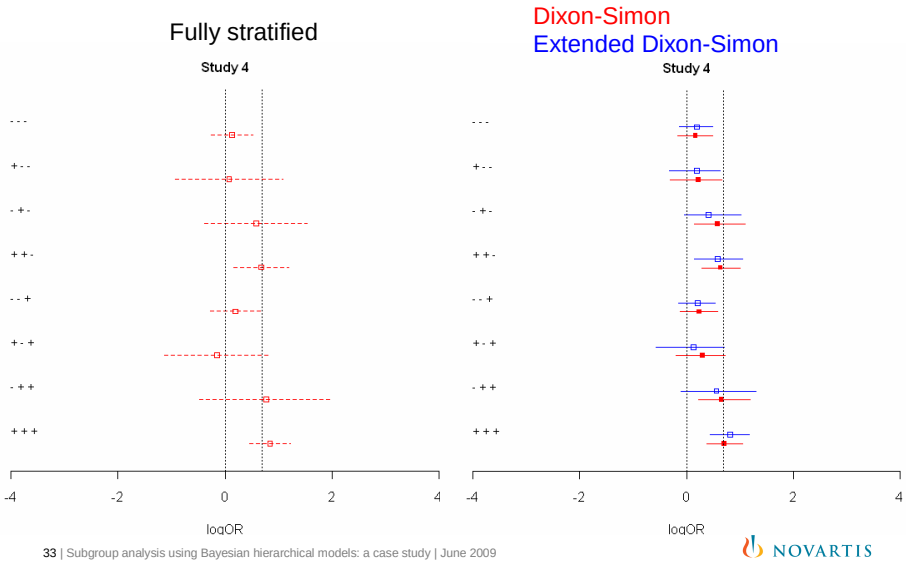
Two subgroup analyses for Study 1



NOVARTIS

Case Study

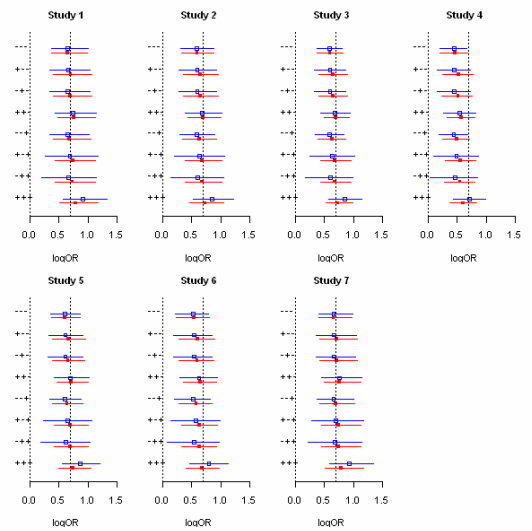
Two subgroup analyses for Study 4



NOVARTIS

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

NOVARTIS

Concluding Remarks

NOVARTIS

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

NOVARTIS

Concluding Remarks

- Further considerations
  - Model diagnostics: residual analyses, posterior predictive checks
  - Model selection (e.g. deviance information criterion DIC, Spiegelhalter et al, JRSS(B), 2002)
  - Recombination of subgroups to larger subgroups
  - Computations: e.g. *WinBUGS*

BUT!

  - Deciding on an analysis after looking at the data is “dangerous, useful, and often done”; Jack Good (*Good Thinking*, 1983)
  - Recommendation: pre-define subgroups and use estimation approach based on shrinkage methods for the analysis
- Manuscript in preparation

NOVARTIS