Subgroup Analysis For Regulatory Purposes: A View From an Industry Statistician

Oliver Keene (GSK)

Acknowledgements / Conflict of Interest

- I am a fulltime GSK employee and own shares in the company
- This presentation includes post-hoc analysis from the DREAM, MENSA and MUSCA studies funded by GSK
  - GSK Study IDs: 112997, 115588, 200862; NCT01000506, NCT01691521, NCT02281318)
- Published as:

  Pavord ID et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Lancet 2012; 380: 651-659.

  Ortega, HG et al. Mepolizumab treatment in patients with severe eosinophilic asthma. NEJM 2014: 371(13), 1198-1207

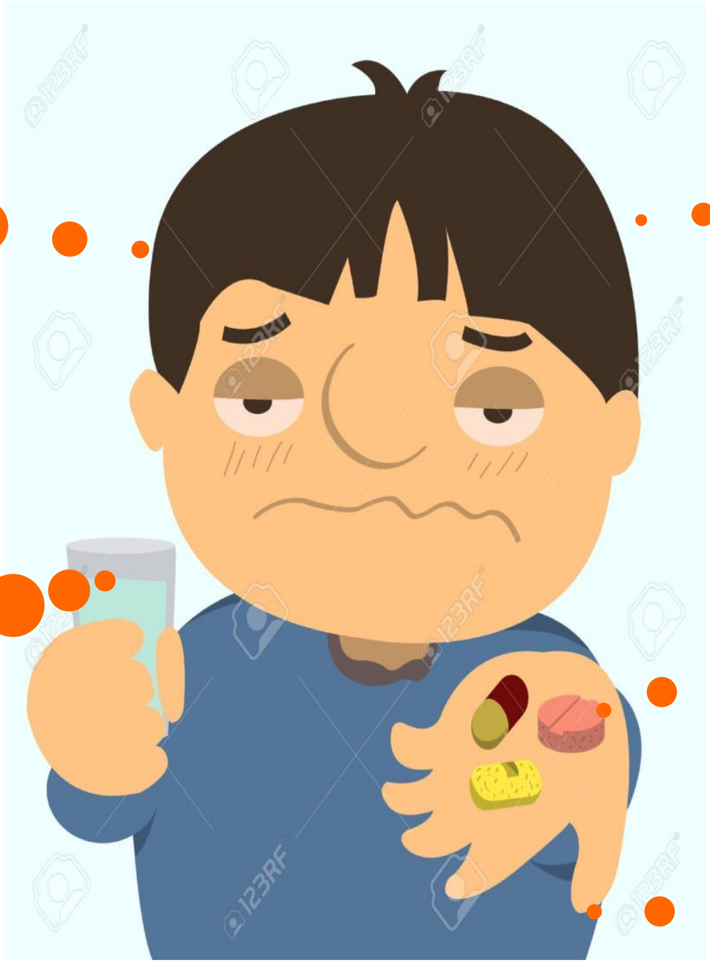
  Chupp GL et al. Efficacy of mepolizumab add-on therapy on health-related quality of life and markers of asthma control in severe eosinophilic asthma (MUSCA): a randomised, double-blind, placebo-controlled, parallel-group, multicentre, phase 3b trial. Lancet Respiratory Medicine 2017; 5:390-400.

Outline

- Multiplicity
- Assessing subgroup effects
  - Scale of measurement
  - Continuous covariates
  - Interaction tests
- Bayesian extrapolation
- Conclusions

Multiplicity

Multiplicity: Typical List of Subgroup Analysis



Age
Sex
Race

Region
Country

Baseline severity Events in past year

Concomitant meds

Blood biomarkers

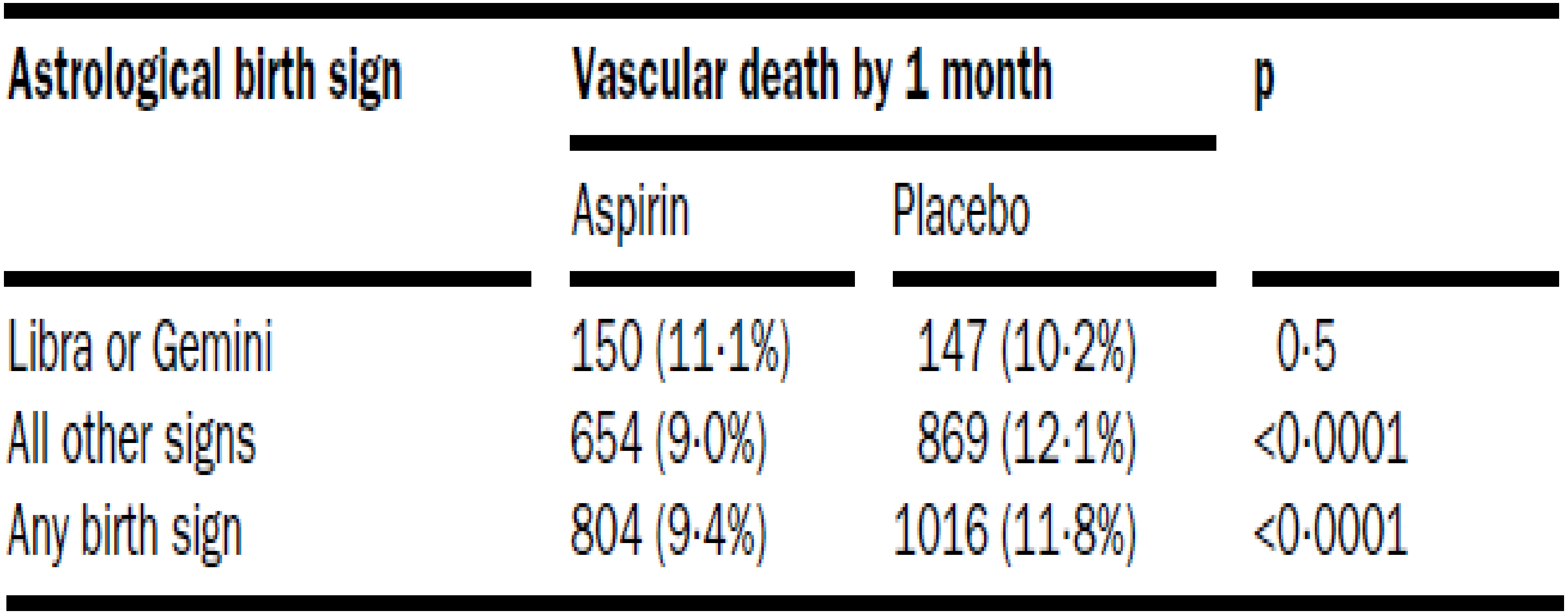
Multiplicity

- Results from analyses are interpreted as the true results for that group of patients
- Subgroup differences in treatment effect can arise by chance
  - Hard to identify what is a true difference
- Single subgroup with 5 levels, equal n, 90% power to detect overall effect*
- No true difference among subgroups
- Probability of observing at least one negative subgroup result = 32%

* Li Z, Chuang-Stein C, Hoseyni C. Drug Inf J. 2007;41(1):47–56

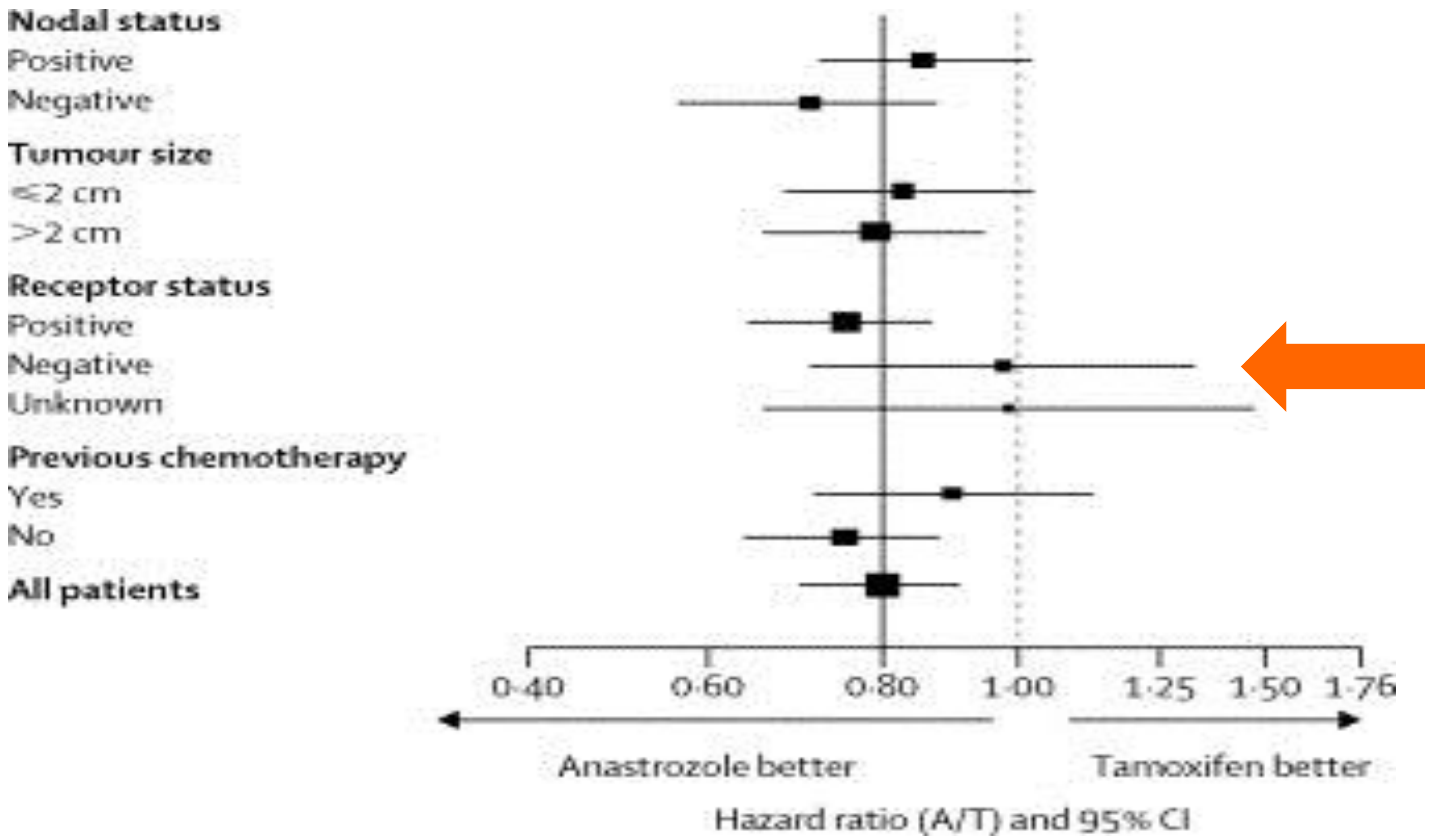
Classic Example: ISIS-2 trial

Trial of aspirin in 17000 subjects



Astrological birth sign	Vascular death by 1 month		p
	Aspirin	Placebo	
Libra or Gemini	150 (11·1%)	147 (10·2%)	0·5
All other signs	654 (9·0%)	869 (12·1%)	<0·0001
Any birth sign	804 (9·4%)	1016 (11·8%)	<0·0001

Example Forest Plot



Does this indicate a lack of effect in negative/unknown receptor state?

Cuzick J. Forest plots and the interpretation of subgroups. Lancet 2005 9;365(9467):1308.

Multiplicity: is the Difference Real?

- Biological plausibility is important
  - Helpful to pre-define this e.g.
    - Differential effect anticipated
    - Plausible but not anticipated
    - Not plausible, hypothesis generating
- Consistency across endpoints (but endpoints typically correlated)
- Replication across two trials
  - If unexpected result is not replicated, then evidence for a true difference is weaker
  - But if no true difference, then 50% chance direction of effect will be the same in the two trials

Design Assumption

Frequent assumption by sponsors

- Patient population is homogeneous
  - Pragmatic approach for sample size determination
  - Expect a consistent treatment effect, anything else due to chance

Alternative assumption:

Treatment effect will vary between subgroups

Burden of proof to establish an effect in each heterogeneous subgroup is with the trial sponsor

Can we Limit the Number of Subgroups?

- Design stage, pre-specification
  - Scientific rationale for heterogeneous effects?
  - Should separate trials be performed?
  - Pre-agreement with regulatory authorities on important subgroups may be helpful
- Need for subgroup analysis is related to the overall patient population
  - Sponsors may identify targeted populations
  - The more homogeneous the population studied, the fewer requirements there should be for subgroup analyses

Assessment of consistency across subgroups

Different Background Rate or Different Treatment Effect?

Events/yr	Placebo	Active	Absolute reduction	Percentage reduction
Baseline				
0	0.8	0.6	0.2	25%
1	1.2	0.9	0.3	25%
2 or more	1.8	1.35	0.45	25%

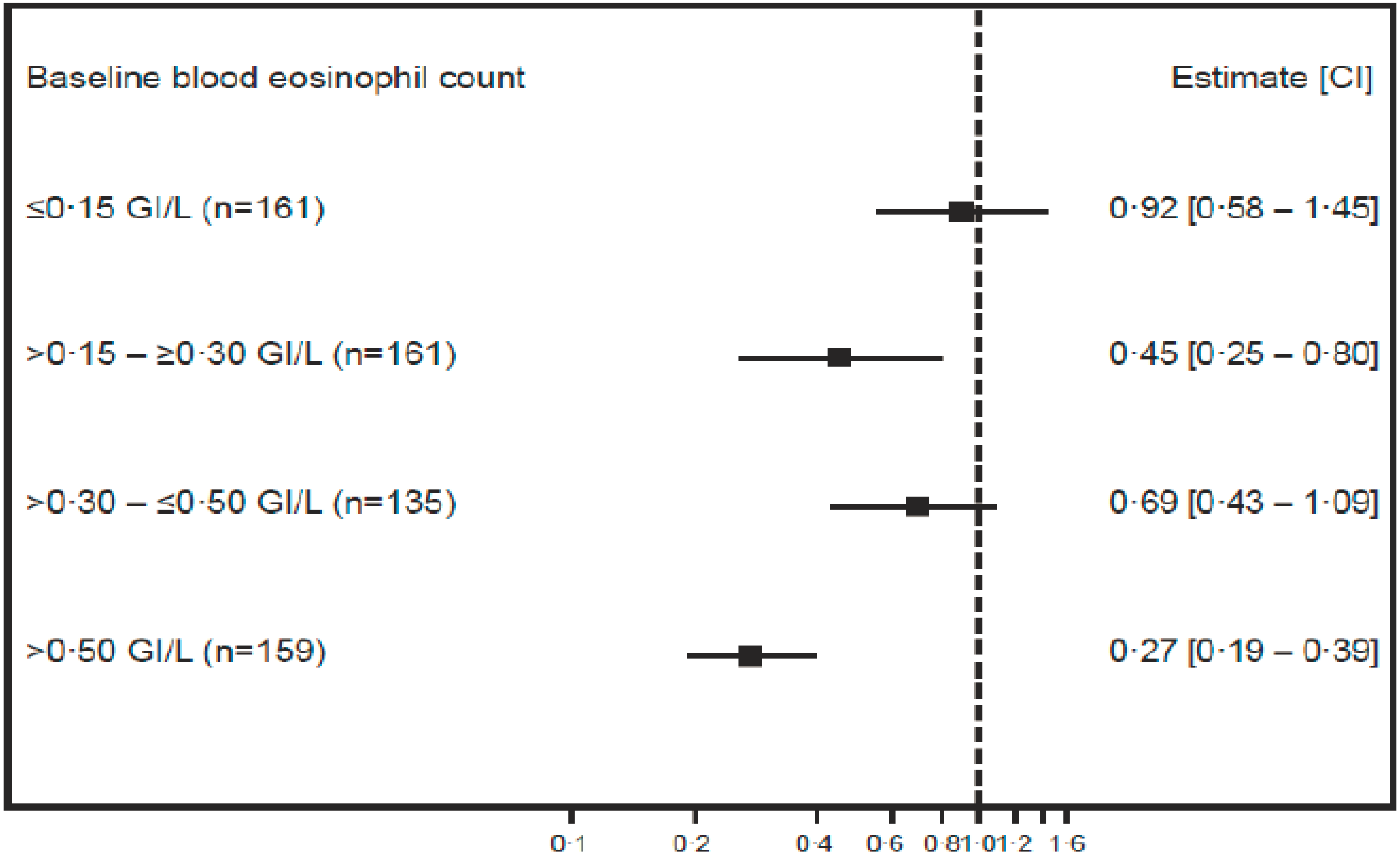
Results are hypothetical and not taken from an actual trial

Continuous not Categorical

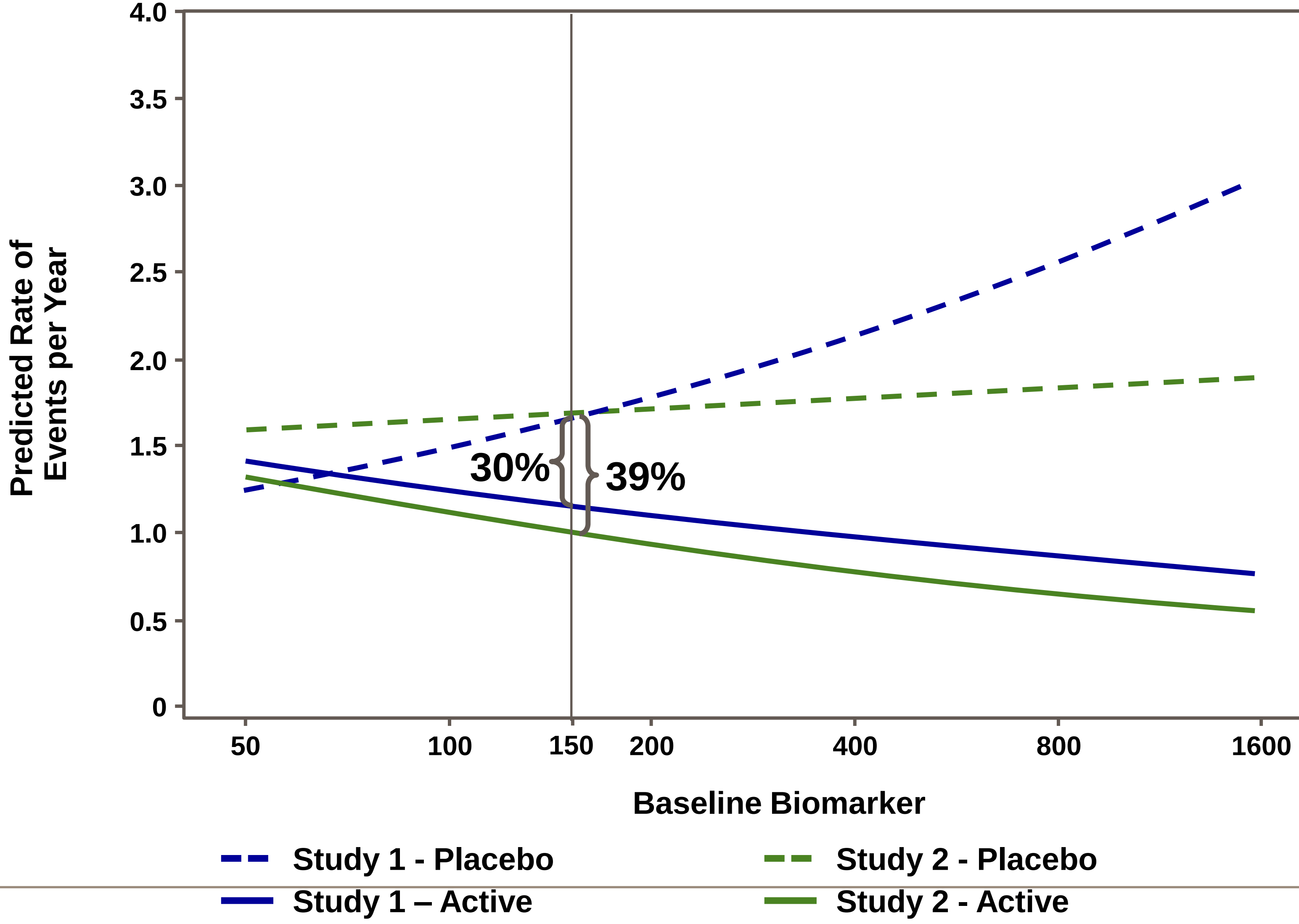
- Typical to classify continuous variable e.g.biomarkers into categories
- Disadvantages:
  - Loss of information
  - Patients close to cutpoint assumed to have very different responses when these are likely to be similar e.g. age 64 vs 65
- Preferable where possible to model relationship between response and continuous covariate

- Example effect of new active treatment vs. baseline levels of a predictive biomarker, assessed in 2 trials

Traditional Presentation: Trial 1, Efficacy by Categories



Predicted Event Rate by Baseline Biomarker: Continuous Scale



Standard Approaches to Consistency

Interaction tests

- Of limited value when investigating subgroup differences
  - Low power to detect heterogeneity
  - Still have 5% or 10% false positive rate
  - Hypothesis testing not appropriate
- Estimates of size of interaction can be helpful to show what differences a trial can reliably estimate

Effect sizes

e.g. Require effect size subgroup to be positive

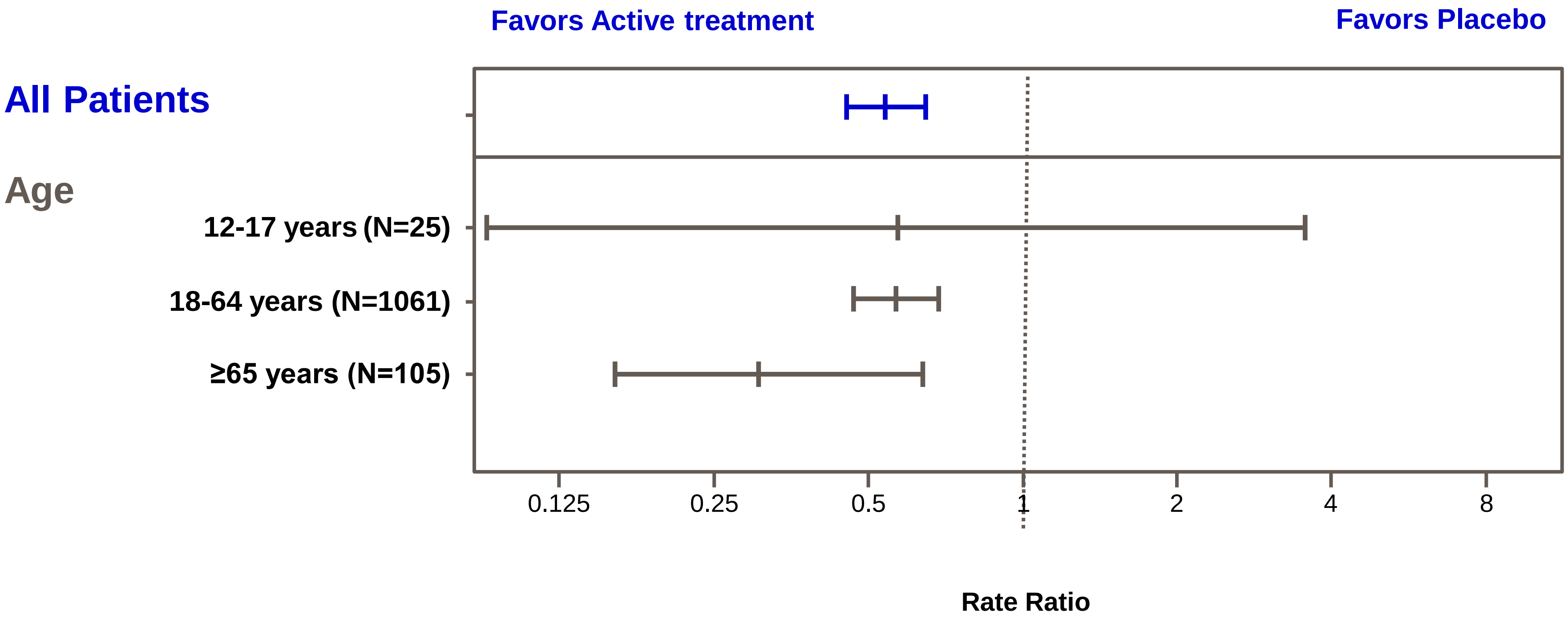
- 50% chance that if the drug has no effect in that subgroup, trial will show a positive effect in the subgroup
- Still high probability of effect reversal by chance if drug actually has desired effect

Example: Bayesian Extrapolation to Adolescent Subgroup

- Severe eosinophilic asthma has late onset and primarily exists in adults
- But some children also suffer (unmet medical need)
- Due to the low incidence, separate clinical efficacy studies not feasible

- Recruitment of phase III trials primarily in adults
  - Two trials recruited adolescent subjects:(aged 12-17)
  - Adults n = 1093, adolescents n=34
  - Can we assess how much belief needed in adult data to infer positive evidence of effect in adolescents?

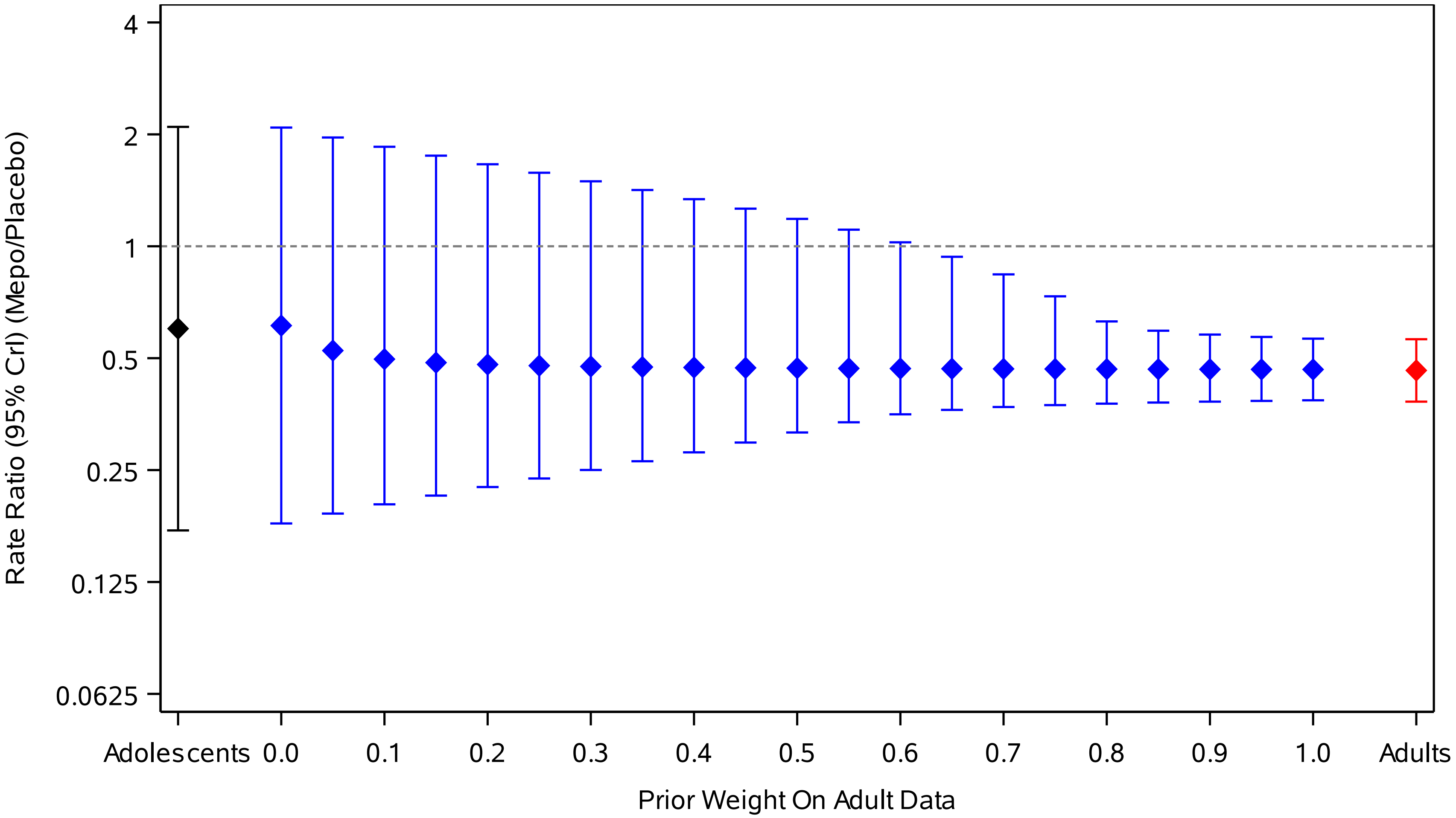
Primary Endpoint by Age group: Trial 1



Bayesian Extrapolation Analysis

- If no strong plausibility for a different effect in a subgroup, then overall trial result is a guide to the effect in that subgroup as well as the estimated effect in the specific subgroup
- Bayesian extrapolation analysis for a subgroup:
  - Construct mixture prior of informative effect in complementary subgroup and uninformative prior
  - Can vary prior weight given to informative prior (analysis updates the weight)
  - – One approach: determine how strong the weight needs to be on informative component for 95% credible interval to exclude no effect (corresponds to one sided p<0.025)
- Provides compromise between assuming effect in subgroup is same as overall effect and using only the data from that subgroup

Posterior Median, 95% CrI against Prior Weight for Adults



Conclusions

Conclusions

- Subgroup analysis is major statistical challenge
  - Hard to identify true effects versus false positives
  - Pre-identification of limited number helpful for interpretation
  - Subgroup analysis should depend on heterogeneity of the population
    - Less requirement when population is targeted
- Difficult to define consistency of effect
  - Modelling of continuous covariate not classification
  - Interaction tests are of doubtful value
- Bayesian extrapolation approaches may be potentially useful

References

JOURNAL OF BIOPHARMACEUTICAL STATISTICS

Special Issue: Subgroup Analysis in Clinical Trials, Volume 24, Number 1, 2014

- *R. Hemmings .* An Overview of Statistical and Regulatory Issues in the Planning, Analysis, and Interpretation of Subgroup Analyses in Confirmatory Clinical Trials.
- *S.-J. Wang and H. M. James Hung.* A Regulatory Perspective on Essential Considerations in Design and Analysis of Subgroups When Correctly Classified
- *A. Koch and T. Framke.* Reliably Basing Conclusions on Subgroups of Randomized Clinical Trials
- ***O. N. Keene and A. D. Garrett.* Subgroups: Time to Go Back to Basic Statistical Principles?**
- *G. G. Koch and T. A. Schwartz.* An Overview of Statistical Planning to Address Subgroups in Confirmatory Clinical Trials