

Subgroup Analysis: A View From an Industry Statistician

Oliver Keene,
GlaxoSmithKline

- **I am a full-time employee of GlaxoSmithKline and I hold shares in the company**
- **The views expressed in this presentation are personal and do not necessarily represent those of GlaxoSmithKline or of the Pharmaceutical Industry in general**

Subgroup Analysis

- Of interest to
 - Regulators
 - Payers
 - Pharma industry
 - Patients
- Aim:
 - Identify patient groups with differential treatment effects
 - Assessment of internal consistency
 - Concern that the response of the “average” patient may not be the response of the patient being treated

Outline

- **Specifying subgroup differences**
 - Scale of measurement
 - Continuous covariates
- **Multiplicity**
- **Design assumptions**
- **Performing subgroup analyses**
 - Assessing consistency of effect

Different Background Rate or Different Treatment Effect?

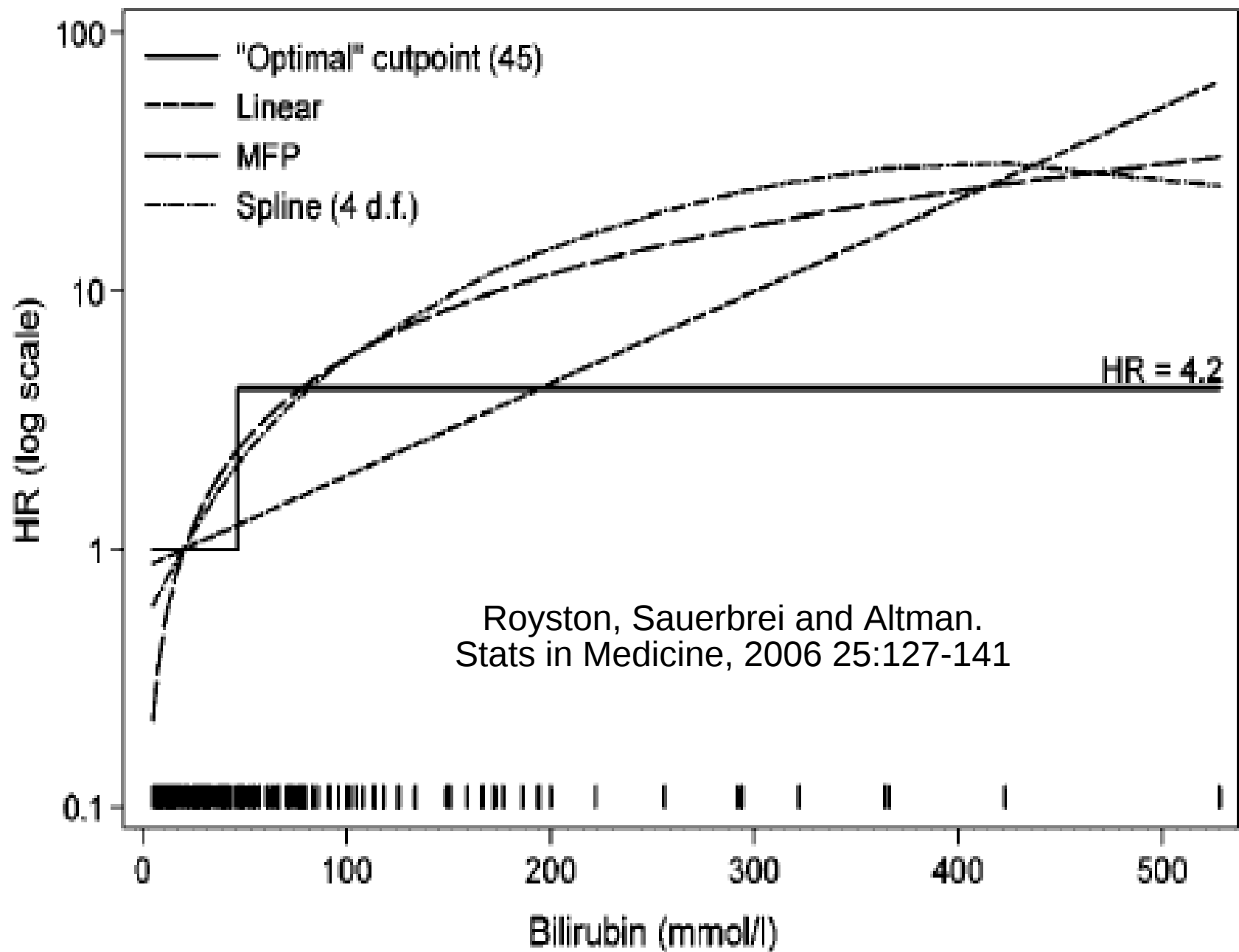
Events/yr	Placebo	Active	Absolute reduction	Percentage reduction
Baseline				
0	0.8	0.6	0.2	25%
1	1.2	0.9	0.3	25%
2 or more	1.8	1.35	0.45	25%

Or Both?

Events/yr	Placebo	Active	Absolute reduction	Percentage reduction
Baseline				
0	0.78	0.64	0.14	19%
1	1.20	0.89	0.31	26%
2 or more	1.75	1.21	0.54	35%

Continuous not Categorical

- Typical to classify continuous variable such as age 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 to model relationship between response and continuous covariate



Multiplicity

- 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 aspirin vs placebo for vascular deaths**
- **Subgroup analysis by star sign**
 - Gemini or Libra: adverse effect of aspirin on mortality
 - Remaining star signs: highly significant effect of aspirin on mortality

ISIS-2. Lancet 1988; 332:349-360

Multiplicity: Typical List of Subgroup Analysis

- Region
- Sex
- Age
- Race
- Baseline severity measure 1
- Baseline severity measure 2
- Clinical events in the previous year
- Baseline medication
- Baseline blood biomarker

Multiplicity: is the Difference Real?

- **Biological plausibility**
- **Pre-definition**
 - Differential effect anticipated
 - Plausible but not anticipated
 - Not plausible, hypothesis generating
- **Consistency across endpoints**
- **Replication across two trials**
 - But meta-analysis can still have subgroup problems
 - More work needed on false positives/false negatives when there are two trials rather than one

Current CHMP Multiplicity Guideline

“A specific claim of a beneficial effect in a specific subgroup requires pre-specification of the corresponding null hypothesis and an appropriate confirmatory analysis strategy.”

“It is highly unlikely that claims based on subgroup analyses would be accepted in the absence of a significant effect in the overall study population.”

“A licence may be restricted if unexplained strong heterogeneity is found in important subpopulations, or if heterogeneity can reasonably be assumed but cannot be sufficiently evaluated for important subpopulations.”

Design Assumption

- **Frequent assumption (by sponsors?) :**
patient population is homogeneous
 - Pragmatic approach for sample size determination
 - Should expect a consistent treatment effect
 - Anything else due to chance
- **Alternative assumption (by regulators?):**
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

Performing Subgroup Analyses – Current Guidelines

ICH E9:

“... should proceed first through the addition of interaction terms to the statistical model in question, complemented by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates.”

CONSORT 2010:

“When evaluating a subgroup the question is ... whether the subgroup treatment effects are significantly different from each other. To determine this, a test of interaction is helpful, although the power for such tests is typically low.”

Simple Interaction Tests

- Tests for interaction 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

Consistency of Effect

- **Measure 1:**

Effect size in each subgroup must at least be positive

- 50% chance that if the drug has no effect in that subgroup, trial will show a positive effect in the subgroup

Consistency of Effect

- **Measure 2:**

Effect size in each subgroup must be at least 50% of overall effect

- Not clear how to apply this on log scales e.g. hazard ratio
- Focus is only on estimate, no account of variability
- 50% chance that if the drug has 50% of overall effect in that subgroup, trial will show a >50% effect in the subgroup

Potential Anomaly?

- Trial 1: overall result 20 units
 - Subgroup 1: 8 units
 - Subgroup 2: 32 units
 - Approval subgroup 2 only?
- Trial 2: overall result 10 units
 - Subgroup 1: 6 units
 - Subgroup 2: 14 units
 - Approval for both subgroups?

Bayesian Shrinkage Estimates

- “Overall trial result is usually a better guide to the effect in the subgroups than the estimated effect in the subgroups” (1)
- Bayesian shrinkage methods (2) combine overall effect with effect in specific subgroup
- Provide compromise between assuming no difference in subgroup and using only the data from that subgroup
 - Implicit prior is that effect in subgroup is same as overall effect

1. Bender R, Lange S. J Clin Epi 2001;54:343-349

2. Jones HE et al. Clinical Trials 2011;8:129-143

3. Simon R. Statist. Med. 2002; 21:2909–2916

Conclusions

- **Subgroup analysis is major statistical challenge**
 - Hard to identify true effects versus false positives
 - Modelling of continuous covariate not classification
- **Pre-identification helpful for interpretation**
 - Is there potential for pre-agreement with regulatory authorities on important subgroups?
- **Subgroup analysis should depend on heterogeneity of the target population**
 - i.e. how broad the inclusion criteria
- **Difficult to define consistency of effect**
 - Interaction tests are of limited value
 - Requirement for each subgroup to show given level of effect is problematic
 - Bayesian approaches may be potentially useful

Conclusions

- **New guideline on subgroup analysis needs to balance**
 - Any increased requirements to show consistency of effect
 - With appropriate consideration of the level of evidence that sponsors are required to provide before a patient in a particular subgroup may receive a new medicine

Acknowledgements

- **PSI expert group on subgroup analysis**
 - Sarah Bujac
 - John Davies
 - Chrissie Fletcher
 - Andrew Garrett
 - Alan Phillips
 - Carol Reid
 - Stephen Sharp