Regulatory considerations for Subgroup Analysis in Clinical Trials

SFDS

Journée nationale B&S 2014

Gautier Paux

27 November 2014

Subgroup analysis in clinical trials

Motivations

- Individuals vary in their response to a treatment
  - Works better for some types of individuals than for other
  - Risk profile of the medicine changes in different individual types
- Considerable challenges for those
  - Designing
  - Analyzing
  - Drawing inferences

Subgroup analysis in clinical trials

Why subgroups are of interest?

- For sponsors and regulators:
  - For which patient group does the medicine show therapeutic efficacy?
  - For which patient is risk-benefit balance favorable?
- For payers:
  - For which patient group does this new medicine represents value for money?
- For prescribers:
  - Which medicine is best for my patient?
- For patients:
  - Do I want to take this medicine?

Outline

- Regulatory considerations for Subgroup Analysis in Clinical Trials
  - Definition and classification
  - Regulatory considerations

Definition and classification

- Subgroup analysis
  - the evaluation of treatment effects with respect to an outcome in subsets of overall trial population defined using patient characteristics at baseline

- Subset definition based on characteristics collected prior to study treatment intervention:
  - Demographic
  - Disease/Medication history
  - Clinical data
  - Genomic data

- Subset definition
  - Dichotomous (e.g. male/female)
  - Categorical (e.g. region)
  - Ordered categorical (e.g. disease score at baseline)
  - Categorized continuous variable (e.g. age class)

- Categorization of subset
  - Pre-specified
  - Justified
  - Sensitivity of the cut-off

- General classification
  - **Exploratory subgroup analysis** focuses on large number of loosely defined patient subgroups
    - Consistency assessment
    - Subgroup identification
  - **Confirmatory subgroup analysis** relies on small number of well defined patient subgroups

- Exploratory subgroup analysis
  - Consistency assessment:

Properties	Consistency assessment
Goal	To evaluate robustness of treatment benefit across multiple subgroups
Number of subgroups	Moderate to large
Scientific rationale	Immaterial
Pre-specification	Not always
Control of Type I error rate	Not needed
Power for testing hypothesis	Inadequately powered

- Exploratory subgroup analysis
  - Subgroup identification:

Properties	Subgroup identification
Goal	To generate hypotheses for further study
Number of subgroups	Probably large
Scientific rationale	Weak or none
Pre-specification	None
Control of Type I error rate	Weak or none
Power for testing hypothesis	Immaterial

- Confirmatory subgroup analysis

Properties	Confirmatory subgroup analysis
Goal	To test hypotheses related to subgroup effects
Number of subgroups	Small (1-2)
Scientific rationale	Strong
Pre-specification	Fully pre-specified
Control of Type I error rate	Mandatory
Power for testing hypothesis	Adequately powered

Subgroup Analysis Classification

Properties	Exploratory subgroup analysis		Confirmatory subgroup analysis
	Consistency assessment	Subgroup identification	
Goal	To evaluate robustness of treatment benefit across multiple subgroups	To generate hypotheses for further study	To test hypotheses related to subgroup effects
Number of subgroups	Moderate to large	Probably large	Small (1-2)
Scientific rationale	Immaterial	Weak or none	Strong
Pre-specification	Not always	None	Fully pre-specified
Control of Type I error rate	Not needed	Weak or none	Mandatory
Power for testing hypothesis	Inadequately powered	Immaterial	Adequately powered

Regulatory considerations

A topic of current interest

2011
- EMA Expert workshop on subgroup analysis

2012
- FDA Enrichment strategies for clinical trials

2014
- February: Draft Guideline on the investigation of subgroups in confirmatory clinical trials (EMA/CHMP)
- August: FDA action plan to enhance the collection and availability of demographic subgroup data
- November: EMA workshop on the investigation of subgroups in confirmatory clinical trials

Active working groups in Europe and USA

- Draft Guideline on the investigation of subgroups in confirmatory CT (EMA/CHMP)
  - Comments made by several organizations in France merged by the SFDS
  - Sent to the EMA by the EFSPI
  - No feedback on the comments yet
  - Discussion on key comments from the public consultation was planned at the EMA workshop in November

- Relevant guidelines
  - [1] EMA – ICH-E9
  - [2] EMA - PTC on multiplicity issues in CT
  - [3] EMA - PTC on adjustment for baseline covariates
  - [4] EMA - Draft Guideline on the investigation of subgroups in confirmatory CT
  - [5] FDA - 21 CFR 314.50
  - [6] FDA - Good Review Practice Statistical Review Template

Key elements

- *[1] "In most cases, however, subgroup or interaction analyses are exploratory and should be clearly identified as such; […]* ***When exploratory, these analyses should be interpreted cautiously****; any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses are unlikely to be accepted.*

- *[2] "Some factors are known to cause heterogeneity of treatment effects […].* ***Analyses of these important subgroups should be a regular part of the evaluation of a clinical study (when relevant), but should usually be considered exploratory****"*

Key elements

- *[2] "However, when a **strong interaction is found that indicates an adverse effect of the treatment in one of the subgroups** and no convincing explanation for this phenomenon is available or other information confirms the likelihood of an interaction then **patients from the respective sub-population may be excluded from the license until additional data are available**."*

- *[3] "If some interactions turn out to be large from a clinical point of view or significant from a statistical point of view, this provides **evidence that the effect of treatment may vary across subgroups**. These findings should be examined carefully; conclusions based on the primary analysis (with no interaction) should be interpreted cautiously and commented on.*

- Key elements [4]
  - **Heterogeneity:** the more heterogeneous the population, the more important the investigation of **consistency** of effects in well-defined subgroups
  - How to judge the **credibility** of the findings?
    - **Biological plausibility:** a clinical and pharmacological judgment
    - **Replication:** of treatment effects from multiple sources of relevant clinical trial data

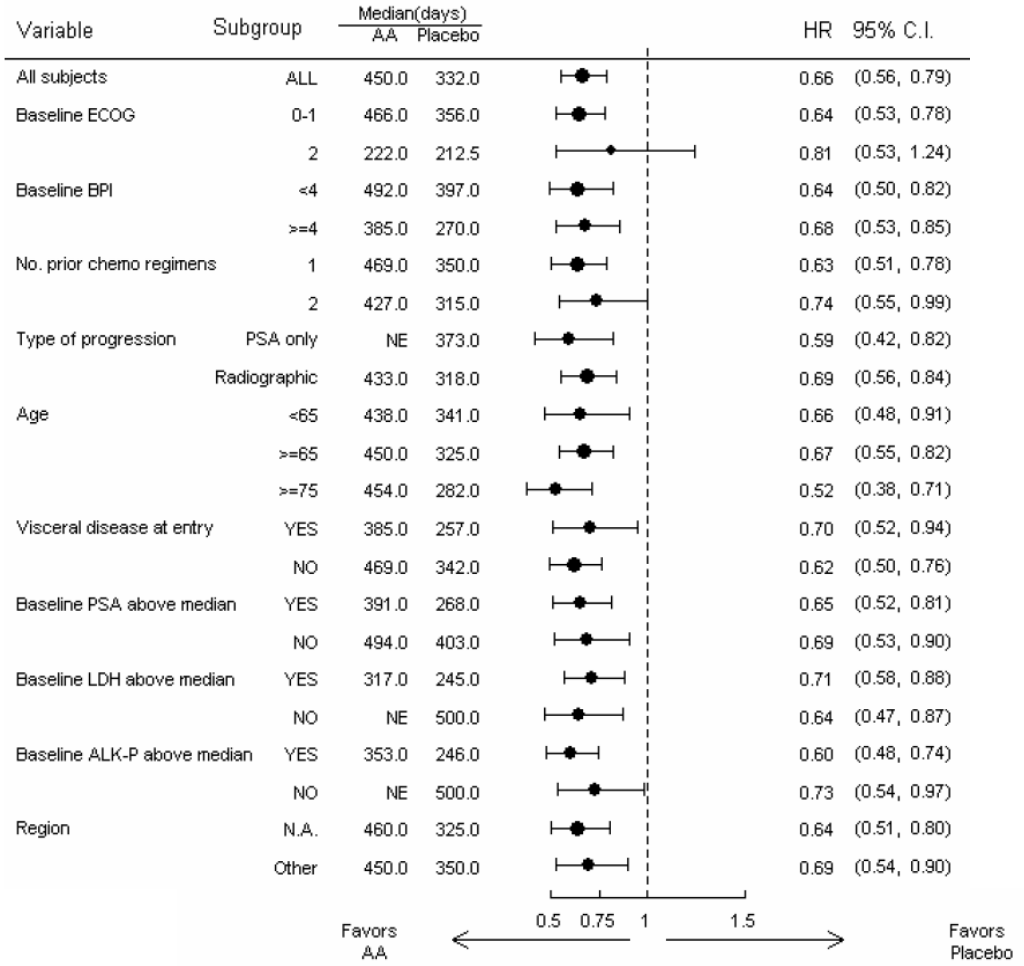
- Key elements [4] – Planning stage
  - **Goal:** to reduce the risk of erroneous conclusions about subset of the population
  - Discussion in the protocol of the expected degree of heterogeneity of the patient population
  - For a factor:
    - At least some biological plausibility or external evidence of the heterogeneous response → **key subgroup**
    - Homogeneity is expected → **exploratory subgroup**
  - Consistency assessment on key subgroups should be planned *a priori*

- Key elements [4]
  - Classification based on 3 scenarios

Scenario	Results of the trial	Analysis
1 (+)	Positive	Consistency assessment
2 (±)	Positive but efficacy or B/R borderline or unconvincing	Identify post-hoc a subgroup: • Exclude due to lack of efficacy • Focus on a subgroup without safety issue
3 (-)	Negative	Identify post-hoc a subgroup who benefit from the treatment

- Key elements [4]
  - Tools for assessing consistency - Methods
    - Test of interaction with estimate, CI and p-value
      - Unpowered
      - Not stratified for most of the factors
      - Qualitative vs. quantitative interaction
    - Subgroup analysis with estimate, CI and p-value
      - Unpowered
      - Not stratified for most of the factors
    - Bayesian shrinkage estimates combining overall and subgroup specific effect (briefly mentioned in the draft guideline)

- Key elements [4]
  - Tools for assessing consistency – Forest Plot



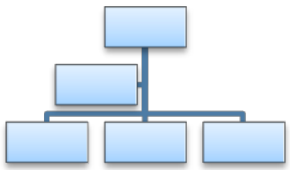
Variable	Subgroup	Median(days) AA	Placebo	HR	95% C.I.
All subjects	ALL	450.0	332.0	0.66	(0.56, 0.79)
Baseline ECOG	0-1	466.0	356.0	0.64	(0.53, 0.78)
	2	222.0	212.5	0.81	(0.53, 1.24)
Baseline BPI	<4	492.0	397.0	0.64	(0.50, 0.82)
	>=4	385.0	270.0	0.68	(0.53, 0.85)
No. prior chemo regimens	1	469.0	350.0	0.63	(0.51, 0.78)
	2	427.0	315.0	0.74	(0.55, 0.99)
Type of progression	PSA only	NE	373.0	0.59	(0.42, 0.82)
	Radiographic	433.0	318.0	0.69	(0.56, 0.84)
Age	<65	438.0	341.0	0.66	(0.48, 0.91)
	>=65	450.0	325.0	0.67	(0.55, 0.82)
	>=75	454.0	282.0	0.52	(0.38, 0.71)
Visceral disease at entry	YES	385.0	257.0	0.70	(0.52, 0.94)
	NO	469.0	342.0	0.62	(0.50, 0.76)
Baseline PSA above median	YES	391.0	268.0	0.65	(0.52, 0.81)
	NO	494.0	403.0	0.69	(0.53, 0.90)
Baseline LDH above median	YES	317.0	245.0	0.71	(0.58, 0.88)
	NO	NE	500.0	0.64	(0.47, 0.87)
Baseline ALK-P above median	YES	353.0	246.0	0.60	(0.48, 0.74)
	NO	NE	500.0	0.73	(0.54, 0.97)
Region	N.A.	460.0	325.0	0.64	(0.51, 0.80)
	Other	450.0	350.0	0.69	(0.54, 0.90)

Source: EMA – Assessment Report for Zytiga (metastatic advanced prostate cancer)

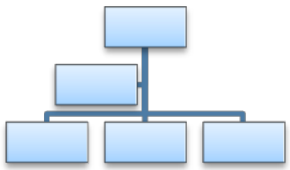
- Key elements [4] - Credibility – Scenario (+)
  - Inconsistent findings in one subgroup considered credible if
    - Biological plausibility and directional consistency and replication

    OR

    - Statistically or clinically extreme results and replication
  - Further supported if :
    - Evidence of treatment-by-covariate interactions across different endpoints
  - Precautionary principle may dictate regulatory action if the replication is unavailable



- Key elements [4] - Credibility – Scenario (±)
  - Level of evidence needed to establish credibility is higher due to problems of **multiplicity** and **data-driven subgroup identification**
  - Required:
    - Subgroup well-defined and clinically relevant entity
    - Pharmacological rationale or mechanistically plausible explanation
    - Estimated treatment effect in the subgroup more pronounced in absolute terms than in the all-randomized population
    - Replication of subgroup findings from other relevant trials



- Key elements [4] - Credibility – Scenario (-)
  - No formal proof of efficacy is possible
  - Provide strong reasons to rescue a failed program:
    - Unmet medical need
    - Trials are of considerable size
    - Same criteria as for scenario 2
    - Clear rationale why properly planned trial has failed (e.g. inclusion criteria)
    - Clear rationale why other studies are unfeasible or unwarranted

- Key elements [5]
  - FDA requires subgroup analyses based on:
    - Age
    - Gender
    - Race
    - Geographic region (if centers outside of the US)

  - *"The **effectiveness data shall be presented by gender, age, and racial subgroups** […]. Effectiveness data from other subgroups of the population of patients treated, when appropriate, […] also shall be presented"*
  - *"The **safety data shall be presented by gender, age, and racial subgroups**. When appropriate, safety data from other subgroups of the population of patients treated also shall be presented"*

- Key elements [6]
  - FDA statistical review and evaluation template

    4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

    4.1 Gender, Race, Age, and Geographic Region

    4.2 Other Special/Subgroup Populations

  - *"The reviewer should describe efficacy (safety) results across subgroups defined **by gender, race, age, and geographic region**. Other subgroups such as those based on baseline characteristics may be included depending on their relevance, representation in the clinical studies, or on the disease being reviewed."*
  - *"The reviewer should include **descriptive statistics** for the defined subgroups. Inferential statistics such as the results of **tests for treatment by subgroup interactions** may also be included. Significant interaction test results should be fully explained, for example by including graphics depicting the results."*

- Relevant guidelines
  - [1] EMA – ICH-E9
  - [2] EMA - PTC on multiplicity issues in CT
  - [3] EMA - PTC on adjustment for baseline covariates
  - [4] EMA - Draft Guideline on the investigation of subgroups in confirmatory CT
  - [5] FDA - Enrichment Strategies for CT

Key elements

- *[1] "In some cases such interactions are anticipated or are of particular prior interest (e.g. geriatrics), and hence **a subgroup analysis, or a statistical model including interactions, is part of the planned confirmatory analysis**."*

- *[2] "Reliable conclusions from subgroup analyses generally require **pre-specification and appropriate statistical analysis strategies**"*

- *[2] "It is **highly unlikely that claims based on subgroup analyses would be accepted in the absence of a significant effect for the overall study population**"*
  - Contradictory with tailored therapy strategies and personalized medicine

- Key elements
  - *[2] "Considerations of **power** expected to be covered in the protocol, and **randomisation would generally be stratified.**"*
  - *[3] "[…] if a substantial treatment by covariate interaction is suspected at the design stage, then **stratified randomisation and/or subgroup analyses should be pre-planned** accordingly. The trial should have **adequate power to detect treatment effects within relevant subgroups**"*
  - *[4] "For a particular factor there is **strong reason to expect a heterogeneous response** to treatment across the different levels of the factor. In this case **separate trials should be usually planned**"*

- Key elements
  - *[5] "[…] the enrichment characteristics used in confirmatory studies should be **measured at baseline**, and patients who are classified as having, or not having, the predictive marker should be **stratified and randomly assigned to treatments** if both subgroups of patients are to be included".*
  - *[5] "[…] the type-I error rate for the study can be shared between **a test conducted using only the enriched subpopulation and a test conducted using the entire population**"*

Key messages

Subgroup Analysis in Clinical Trials

- Key messages
  - Homogeneity of treatment effect is rarely plausible and subgroup analyses should depend on heterogeneity of the target population
  - Pre-identification of subgroups is helpful for interpretation
  - Difficult to define consistency of effect
  - Biological plausibility and replication are the most important concepts in credibility of subgroup findings
  - Regulatory agencies are aware of the pitfalls of the subgroup analyses

Questions / Answers

References

Subgroup Analysis in Clinical Trials

References

- Grouin J. M., Coste M. and Lewis J. Subgroup Analyses in Randomized Clinical Trials: Statistical and Regulatory Issues. Journal of Biopharmaceutical Statistics. 2006
- Hemmings R. An overview of statistical and regulatory issues in the planning, analysis and interpretation of subgroup analyses in confirmatory clinical trials. Journal of Biopharmaceutical Statistics. 2014
- John H. E. *et al*. Bayesian models for subgroup analysis in clinical trials. Clinical trials. 2011
- Lipsky A. R. *et al*. The Importance of "Shrinkage" in Subgroup Analyses. Ann Emerg Med. 2010
- Varadhan R. *et al.* A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes-research. Journal of clinical epidemiology . 2013
- European Medicines Agency. (1998). ICH E9 Statistical Principles of Clinical Trials (CPMP/ICH/363/96)
- European Medicines Agency. (2002). Points to consider on multiplicity issues in clinical trials. CPMP/EWP/908/99

Subgroup Analysis in Clinical Trials

References

- European Medicines Agency. (2003). Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99.
- European Medicines Agency. (2014). Draft guideline on the investigation of subgroups in confirmatory clinical trials. CHMP/539146/2013.
- Food and Drug Administration. (2012). Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products.
- Food and Drug Administration. (2012). Good Review Practice: Statistical Review Template.
- Food and Drug Administration. (2014). Fda Action Plan To Enhance The Collection And Availability Of Demographic Subgroup Data.
- Dmitrienko A. Subgroup Analysis in Clinical Trials. *Formation SFdS B&S*. Vol. 1 .2014.

Back-up

- A treatment-covariate exists when the treatment effect is not the same for all value of the covariate

- Quantitative interaction
  - Treatment effects in the same direction but of different magnitude in some subgroups

- Qualitative interaction
  - Treatment effects in opposite direction

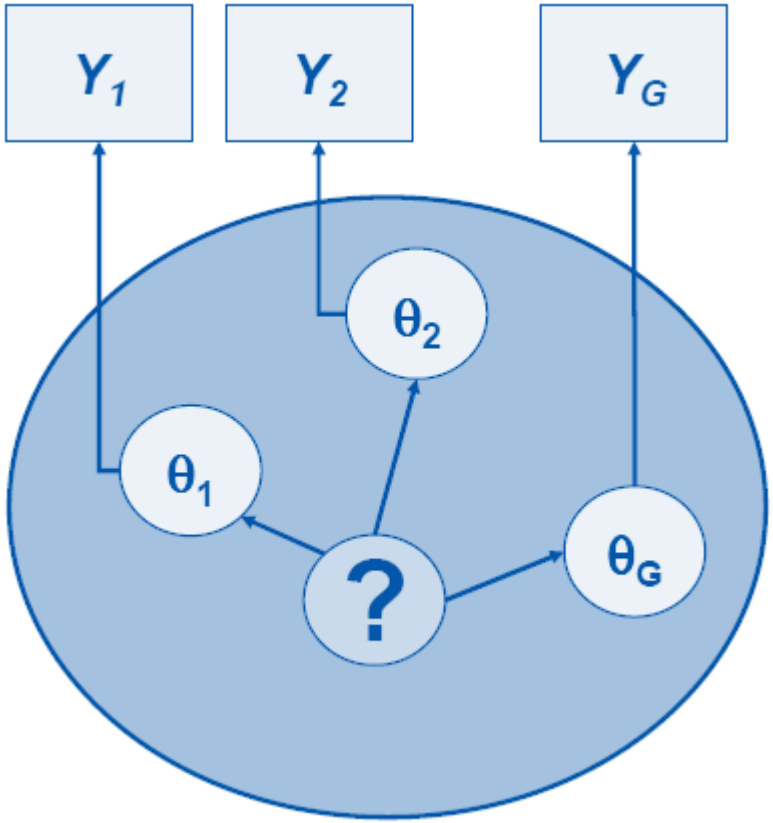
- Treatment X (0 for control, 1 for experimental)
- Covariate Z (e.g Z=0 for female, 1 for male)
- Outcome $Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$

	Control	Experimental	Trt effect
Female	β_0	$\beta_0 + \beta_1$	β_1
Male	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta 1 + \beta_3$
Gender effet	β_2	$\beta_2 + \beta_3$	

- Bayesian shrinkage estimates
  - 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 θ_1, θ_2,…, θ_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
    - modified point estimates
    - generally shorter confidence intervals

Source: Subgroup analysis using Bayesian hierarchical models: a case study - Neuenschwander

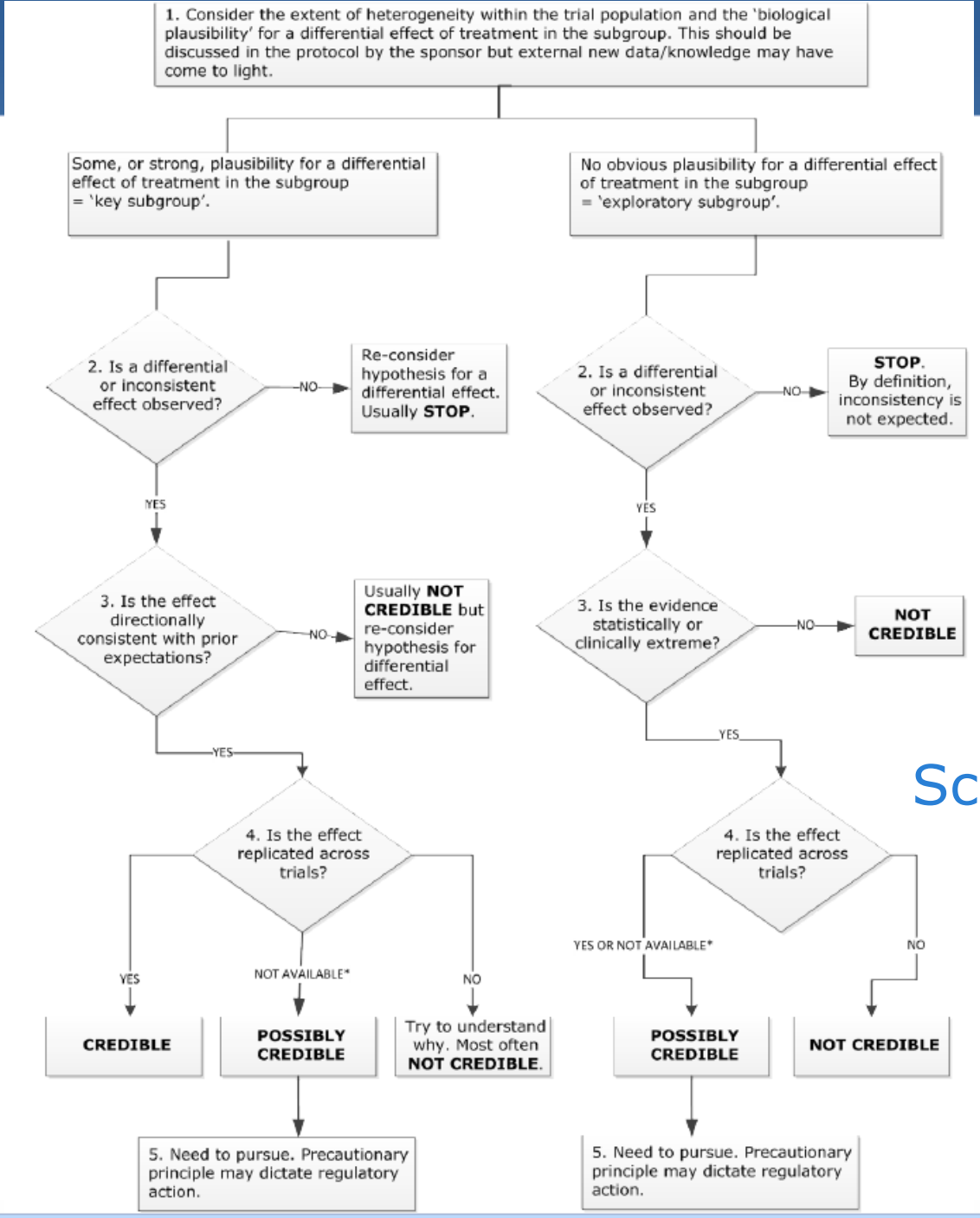
- Shrinkage
  - Y_1, …, Y_G data from G subgroups
  - θ_1, …, θ_G effects
  - ? Unknown Relationship/Similarity
    - From the same effect
    - To very different effect



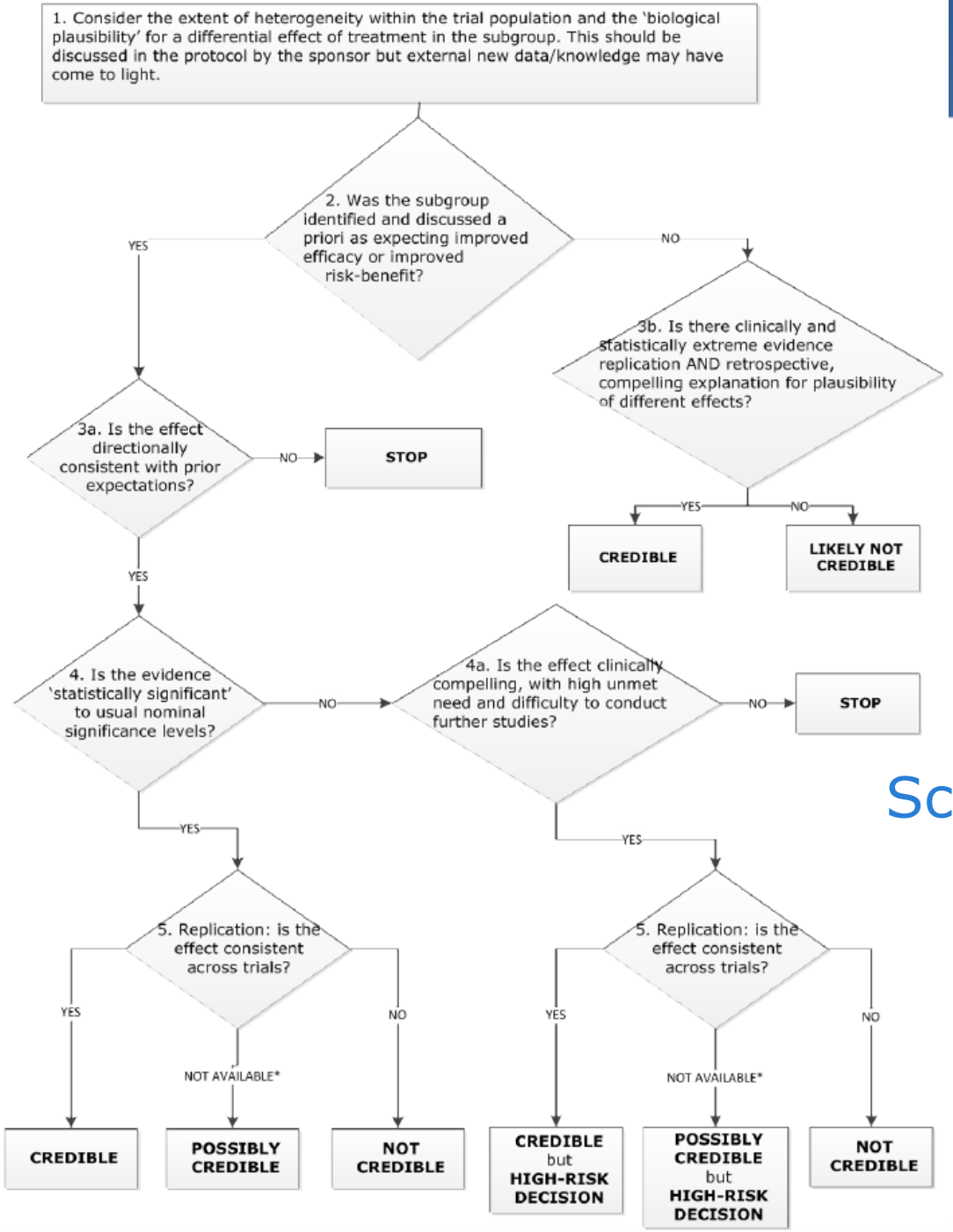
Source: Subgroup analysis using Bayesian hierarchical models: a case study - Neuenschwander

- The simplest model
  - G subgroups with $\theta_1, \ldots, \theta_G$ effects
  - Why shrinkage?
    - Estimates are typically more spread out than true effects $\theta_1, \ldots, \theta_G$
    - Extreme stratified subgroups estimates are typically too extreme
  - Simple shrinkage for subgroup analyses
    - $Y_g \sim N(\mu_g, s_g^2)$, g=1,…,G
    - $\theta_1, \ldots, \theta_G \sim N(\mu, \tau^2)$
  - Inference
    - Classical random-effect analyses
    - Empirical Bayes
    - Fully Bayesian (require prior specification for μ and τ)

Source: Subgroup analysis using Bayesian hierarchical models: a case study - Neuenschwander


1. Consider the extent of heterogeneity within the trial population and the 'biological plausibility' for a differential effect of treatment in the subgroup. This should be discussed in the protocol by the sponsor but external new data/knowledge may have come to light.
Some, or strong, plausibility for a differential effect of treatment in the subgroup = 'key subgroup'.
No obvious plausibility for a differential effect of treatment in the subgroup = 'exploratory subgroup'.
2. Is a differential or inconsistent effect observed?
NO
Re-consider hypothesis for a differential effect. Usually STOP.
YES
3. Is the effect directionally consistent with prior expectations?
NO
Usually NOT CREDIBLE but re-consider hypothesis for differential effect.
YES
4. Is the effect replicated across trials?
YES
CREDIBLE
NOT AVAILABLE*
POSSIBLY CREDIBLE
NO
Try to understand why. Most often NOT CREDIBLE.
5. Need to pursue. Precautionary principle may dictate regulatory action.
2. Is a differential or inconsistent effect observed?
NO
STOP. By definition, inconsistency is not expected.
YES
3. Is the evidence statistically or clinically extreme?
NO
NOT CREDIBLE
YES
4. Is the effect replicated across trials?
YES OR NOT AVAILABLE*
POSSIBLY CREDIBLE
NO
NOT CREDIBLE
5. Need to pursue. Precautionary principle may dictate regulatory action.

Scenario 1


1. Consider the extent of heterogeneity within the trial population and the 'biological plausibility' for a differential effect of treatment in the subgroup. This should be discussed in the protocol by the sponsor but external new data/knowledge may have come to light.
2. Was the subgroup identified and discussed a priori as expecting improved efficacy or improved risk-benefit?
YES
NO
3a. Is the effect directionally consistent with prior expectations?
NO
STOP
YES
3b. Is there clinically and statistically extreme evidence replication AND retrospective, compelling explanation for plausibility of different effects?
YES
NO
CREDIBLE
LIKELY NOT CREDIBLE
4. Is the evidence 'statistically significant' to usual nominal significance levels?
NO
4a. Is the effect clinically compelling, with high unmet need and difficulty to conduct further studies?
NO
STOP
YES
YES
5. Replication: is the effect consistent across trials?
YES
NOT AVAILABLE*
NO
CREDIBLE
POSSIBLY CREDIBLE
NOT CREDIBLE
5. Replication: is the effect consistent across trials?
YES
NOT AVAILABLE*
NO
CREDIBLE but HIGH-RISK DECISION
POSSIBLY CREDIBLE but HIGH-RISK DECISION
NOT CREDIBLE

Scenario 2