Issues Related to Subgroup Analysis and Possible Improvement

Lu Cui, Shufang Liu

AbbVie Inc.



FDA/Industry Workshop 2015
Sep. 16-18, 2015
Washington D.C.

abbvie

Author Disclosure

- All authors are employees of AbbVie Inc.
- The presentation was sponsored by AbbVie. AbbVie contributed to the developing, writing, reviewing and approving the publication. The presentation contains no AbbVie data.

Abstract

Confirmatory randomized clinical trials are designed to provide definitive information on the efficacy and safety of a new drug. While the outcome based on the overall population is the basis for the final conclusion, sound subgroup outcomes may provide additional insights of the drug effects and information for possible development of personalized treatment strategies. This presentation is to revisit issues related to subgroup analysis, to highlight pitfalls of potential biased interpretations and to explore ways for improvements. For the latter, the use of intensive stratification to reduce the chance of biased subgroup findings is discussed with simulation results to illustrate the idea.

Outline

- Subgroup analysis: usage and issues
- Factors affecting the quality and interpretation of subgroup analysis
- Improvement via intensive (but not extensive) stratification
- Conclusion

Subgroup analysis: usage and issues

- Subgroup analyses are widely used in medical researches and the results can be reported in medical publications

- A recent survey of a total 221 publications (Zhang et al, J. of Clinical Oncology, 2015) with data collected from 184,500 patients in RCT showed 188 of them (85%) reported subgroup analyses. Out of those, 144 (78%) reported analyses involving at least six subgroups

- In recent years, subgroup analyses are gaining more attentions partly because of the progress of (e.g. biomarker) technology to help identify potentially more responsive patient populations

Subgroup analysis: usage and issues

- Under regulatory setting, subgroup analyses are traditionally performed in confirmatory clinical trials for regulatory submissions (age<65 / age>=65, males / females, race, and others) and results can be displayed in drug labeling

- Consistent subgroup findings in line with the primary findings may be viewed as supportive evidences of drug efficacy and safety though the results solely based on subgroup analysis, in general, cannot be used to support NDA approvals

- Inconsistent subgroup findings can lead to extra cautions on the intended claims, and may have implications on the drug approvability and labeling

Subgroup analysis: usage and issues

- . Efforts are made for more efficient trials and better targeted treatments (personalized medicine). FDA Guidance for industry on enrichment strategy for clinical trials (2012) contains many examples of subgroup based enrichment designs

Issues with subgroup analysis is always its believability

- Subgroup analysis results have been long viewed suspicious which are good for hypothesis generating but not for confirmation

'Subgroup analyses in clinical trials: fun to look – but don't believe them' (Peter Sleight, Current Control Trials Cardiovascular Med., 2000)

Subgroup analysis: usage and issues

- "Subgroup analyses are an important part of the analysis of a comparative clinical trial. However, they are commonly over-interpreted and **can lead to further research that is misguided or, worse, to suboptimal patient care**" (S. Lagakos, New England Journal of Medicine, 2006)

- "It is widely understood that **subgroup analyses needed to be interpreted with caution** …. An incautious review of subgroup analyses can result in unreliable inferences and, consequently, to poor decisions from the clinical trial sponsor or regulator" (EMA Guideline on Investigation of Subgroups in Confirmatory Clinical Trials, 2014)

Subgroup analysis: usage and issues

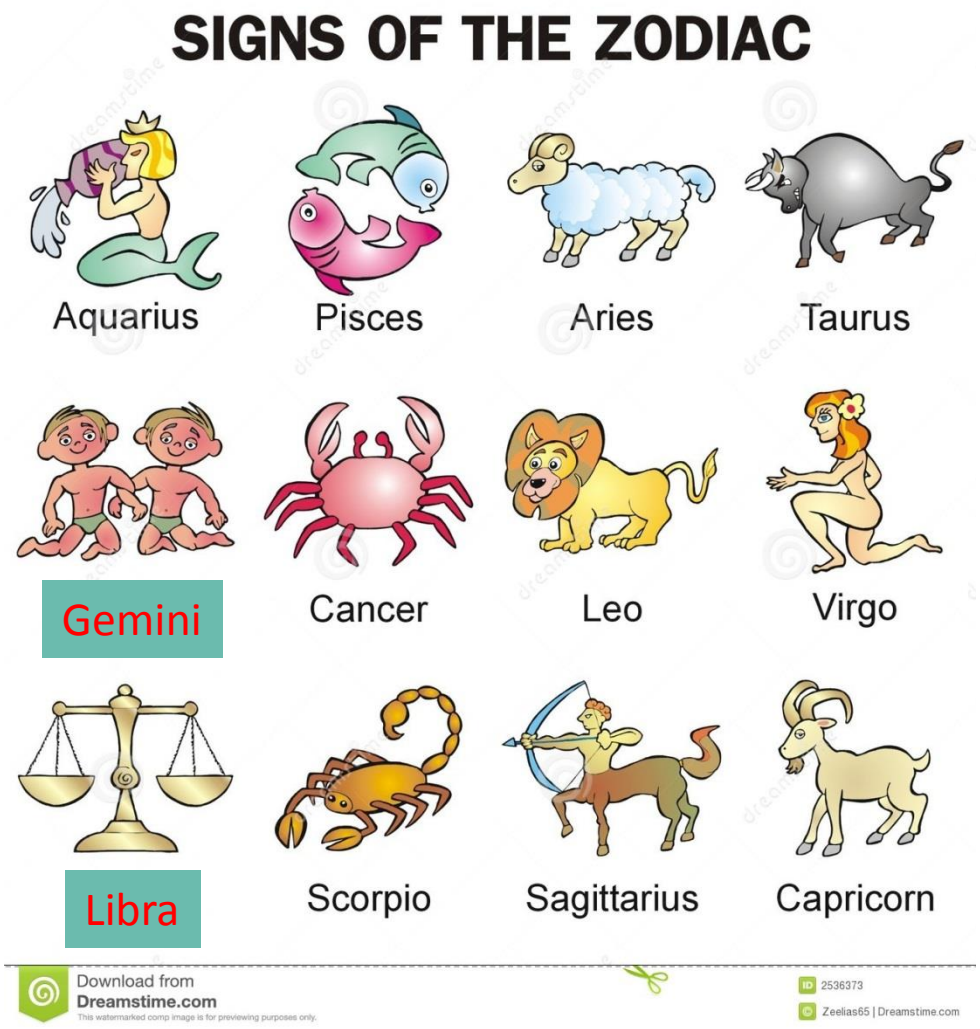
Example 1.

ISIS-2 Trial (1988): patients randomized in a very large Second International Study of Infarction Survival trial (n=17,187)

Study arms: Aspirin vs. placebo
Indication: acute myocardial infarction
Endpoint: 1-month survival

Overall result: significant beneficial effect of Aspirin on survival (804 deaths in Aspirin vs. 1016 deaths in placebo, $p<0.00001$)

Subgroup analysis: usage and issues



- The subgroup analysis in subjects with astrologic sign of Libra or Gemini is not significant

  Aspirin: 150 events vs.
  Placebo: 147

- Should one believe this subgroup finding or not? Why?

Subgroup analysis: usage and issues

Example 2.

- ECST Trial (Rothwell, 2005)
- N=3024 subjects with carotid stenosis, randomized at 3:2, surgery vs. control
- Primary eff. outcome is death
- Believable? Why?

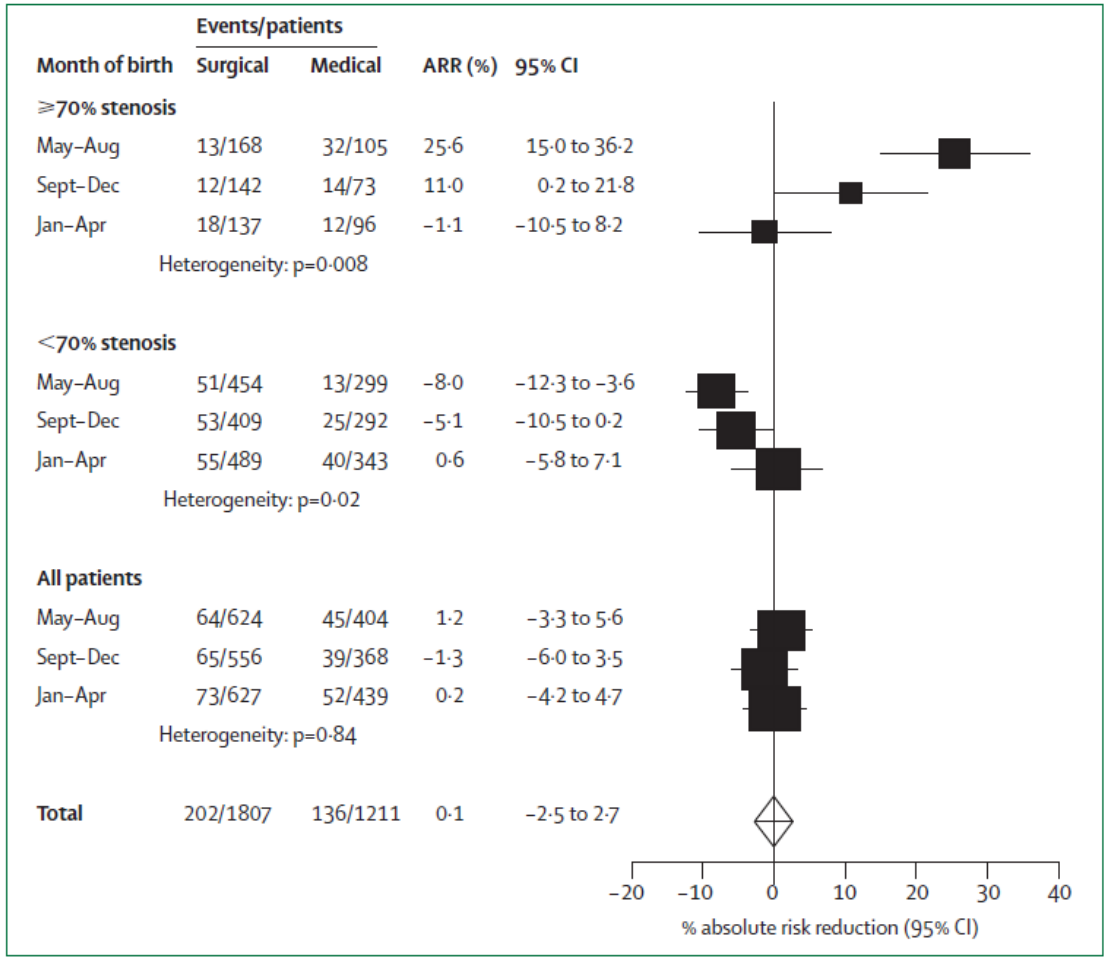


Figure 4: Effect of carotid endarterectomy in ECST[126] according to month of birth in three 4 month periods in patients with ≥70%, <70%, and all degrees of symptomatic stenosis

Subgroup analysis: usage and issues

- No treatment effect was shown in the overall population

- No treatment difference was shown in subgroups by birth-month subgroups

- The treatment effect went to opposite directions in sub-subgroup analysis (Birth-month by % of stenosis)

- ‘Sponsors and regulatory decision makers are put in a difficult situation: whether to accept an assumption of homogeneity and disregard extreme … findings in subgroups, or whether to anticipate some heterogeneity and … attempt to use the results of subgroup analyses as one piece of evidence to inform decision making’ (EMA Guideline, 2014)

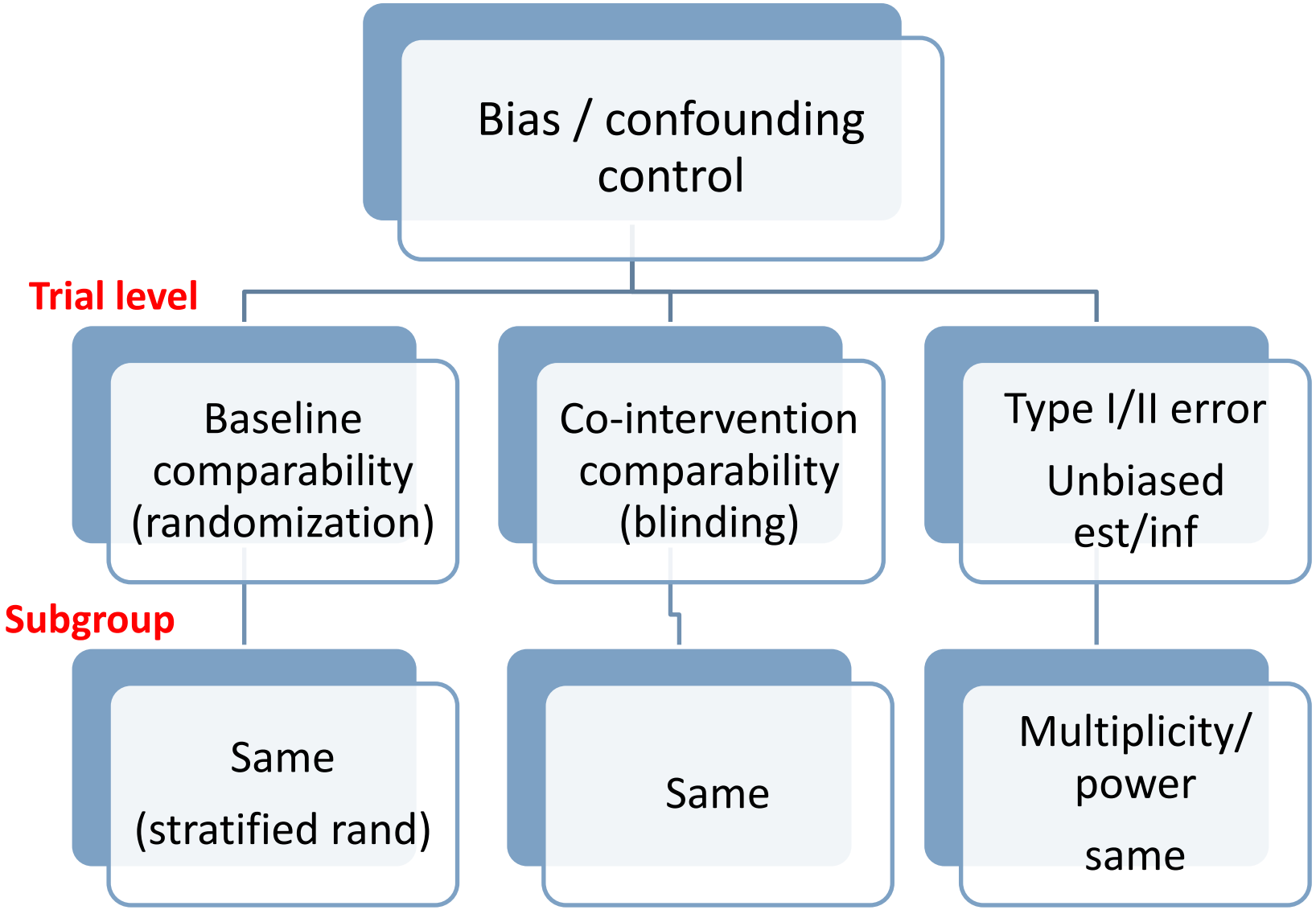
Factors affecting the quality/interpretation of subgroup analysis

- Anna Karenina Principle: 'Happy families are all alike; every unhappy family is unhappy in its own way' (*Anna Karenina*, Leo Tolstoy)

- A clinical trial (subgroup analysis) version:
Good clinical trials (subgroup analyses) are alike; every problematic trial (subgroup analysis) is problematic in its own way

- Instead of figuring out how a subgroup analysis is wrong, we focus on how to make a subgroup analysis good up front

Factors affecting the quality/interpretation of subgroup analysis



Factors affecting the quality/interpretation of subgroup analysis

Violation and impact

	Incomparable Baseline factors	No multiplicity adjustment	Low power
Bias detection	Difficult	Possible	Possible
Bias direction	Unknown	known	known
Bias correction	Difficult	Possible	Possible
Believability	Unknown	Possible	Possible

Factors affecting the quality/interpretation of subgroup analysis

The most vicious issue is likely group incompatibility w.r.t. observable or unobservable influential factors

	Incomparable Baseline factors	No multiplicity adjustment	Low power
Bias detection	Difficult	Possible	Possible
Bias direction	Unknown	known	known
Bias correction	Difficult	Possible	Possible
Believability	Unknown	Possible	Possible

Factors affecting the quality/interpretation of subgroup analysis

- The biases introduced by group incompatibility (involving observed and **non-observed** factors) is hard to detect and hard to correct at the end of the trial

- The direction of impact is unknown

- There is no effective way to remove potential biases since one may not know if there are biases and which covariates should be used to adjust

- The issues associated with multiplicity and power are more benign since in such cases one may be not completely in the dark

Factors affecting the quality/interpretation of subgroup analysis

Some often seen mistakes in interpreting subgroup findings

- Overly believe whatever is seen and superficial post-hoc biological explanations

- Overly pay attentions to consistency but ignore revealing signals from a few

- Overly believe that subgroup analysis is OK as long as it is pre-specified

- Believe the results w/o checking if the analyses satisfied aforementioned criteria to be good

Factors affecting the quality/interpretation of subgroup analysis

- **Quality of a subgroup analysis after all depends on its inherent characteristics w.r.t. group comparability, multiplicity, and power**

- **One way to have fundamental improvement is stratification**

Improve subgroup analysis via intensive stratification

- Our proposal is to consider intensive stratification with 'as many as' potential influential factors as stratification factors to achieve a better baseline group comparability for the overall, subgroup, and sub-subgroup analysis to minimize potential biases due to group incomparability

- Intensive means 'as many as possible' without degenerating into un-stratified randomization (extensive)

- Two schools about stratification: (a) it is unnecessary especially in large trials where complete randomization is fairly efficient, and also covariate adjustment can help (Peto et al); (b) balancing provides a more efficient comparison of treatment and more importantly trials with balanced prognostic factors are far more convincing than sophisticated covariate analyses alone (Begg and Iglewicz)

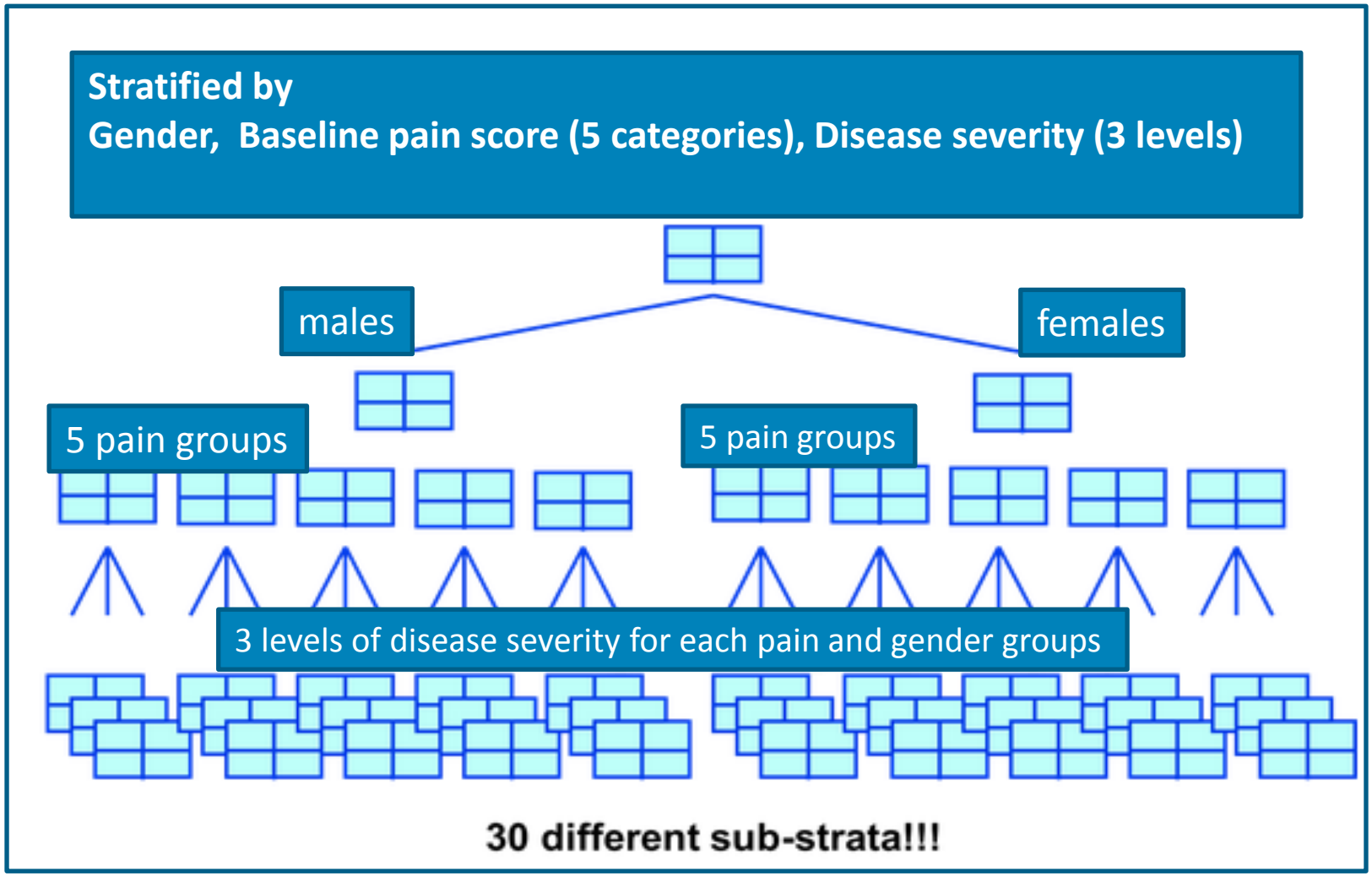
Improve subgroup analysis via intensive stratification

• We takes a position with the elements of (a) and (b): advocating intensive stratification and encourage covariate adjustment

• How intensive should depend on individual studies. It is the trial statistician's job to determine the intensity of the stratification based on the trial requirements and simulation outcomes

• The relevant factors include the total sample size of the trial, the number of treatment groups, the number of influential baseline factors and the number of their possible combinations, the randomization block size and etc.

Improve subgroup analysis via intensive stratification



Improve subgroup analysis via intensive stratification

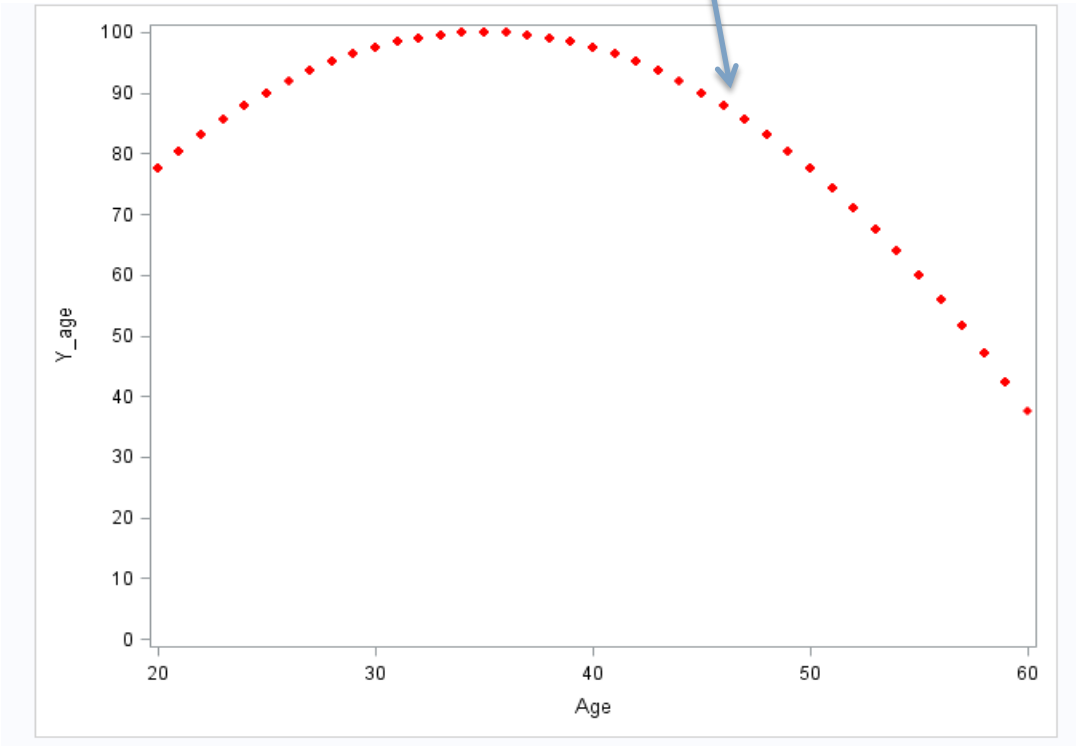
Simulation

- Data is generated based on the model with strong age effect

$$Y = a\,TREAT + 0.1\,(1000\text{-}(AGE - 35)^2) + 10\text{Gender} + \varepsilon$$

H0: a= 0.0

H1: a =5.6



Improve subgroup analysis via intensive stratification

N=200 subjects, gender (male and female) and age (10 categories from 21-25, 26-29,…, 57-60) =>10 subjects per stratum, block sz=2

Effect of stratification (overall population)

Analysis	Estimate	Type I ER	Estimate	Power	SE
Unstratified randomization					
Age correctly adjusted	0.0212	0.0522	5.6212	0.9754	1.42
Adjusted with linear age	0.0085	0.0468	5.6085	0.7112	2.23
No adjustment for age	2.3937	0.1212	7.9937	0.7717	2.96
Stratified randomization	Gender (male, female) by age (10 categories)				
Age correctly adjusted	-0.0075	0.0469	5.6075	0.9771	1.41
Adjusted with linear age	-0.0015	0.0005	5.5015	0.7975	2.21
No adjustment for age	-0.0051	0.0036	5.5949	0.5474	2.75

Improve subgroup analysis via intensive stratification

Effect of stratification (subgroup of females)

Analysis	Estimate	Type I ER	Estimate	Power	SE
Unstratified randomization					
Age correctly adjusted	0.0317	0.0545	5.6317	0.7879	
Adjusted with linear age	0.0234	0.0489	5.6234	0.4162	
No adjustment for age	2.4342	0.0814	8.0342	0.4701	
Stratified randomization	Gender (male, female) by age (10 categories)				
Age correctly adjusted	0.0146	0.0528	5.5994	0.7921	
Adjusted with linear age	0.0029	0.0040	5.6022	0.3834	
No adjustment for age	0.0028	0.0006	5.6019	0.1595	

Improve subgroup analysis via intensive stratification

- Finer stratifications yield similar results.

- Un-stratified randomization may introduce significant bias affecting the parameter estimate and hypothesis testing if no covariate adjustment is made

- Covariate adjustment based on the correct model of the dependency of the outcome on the influential covariate(s) can help to remove the bias but guessing the model correctly is difficult in the practice if not impossible

- Adjustment for the influential covariate(s) based on an incorrectly assumed model may still help to remove the bias but how much is removed generally is unknown

Improve subgroup analysis via intensive stratification

- Stratified randomization gives unbiased parameter estimate and unbiased or conservative inference regardless the use of covariate adjustment

- With the stratified randomization, adjustment for the influential covariate(s) can help to increase the precision of the estimate and increase the power of the statistical testing

- Fundamentally, with the use of the intensive stratification, there will be less concern on the potential bias due to group incomparability in the overall, subgroup and sub-subgroup analyses

Improve subgroup analysis via intensive stratification

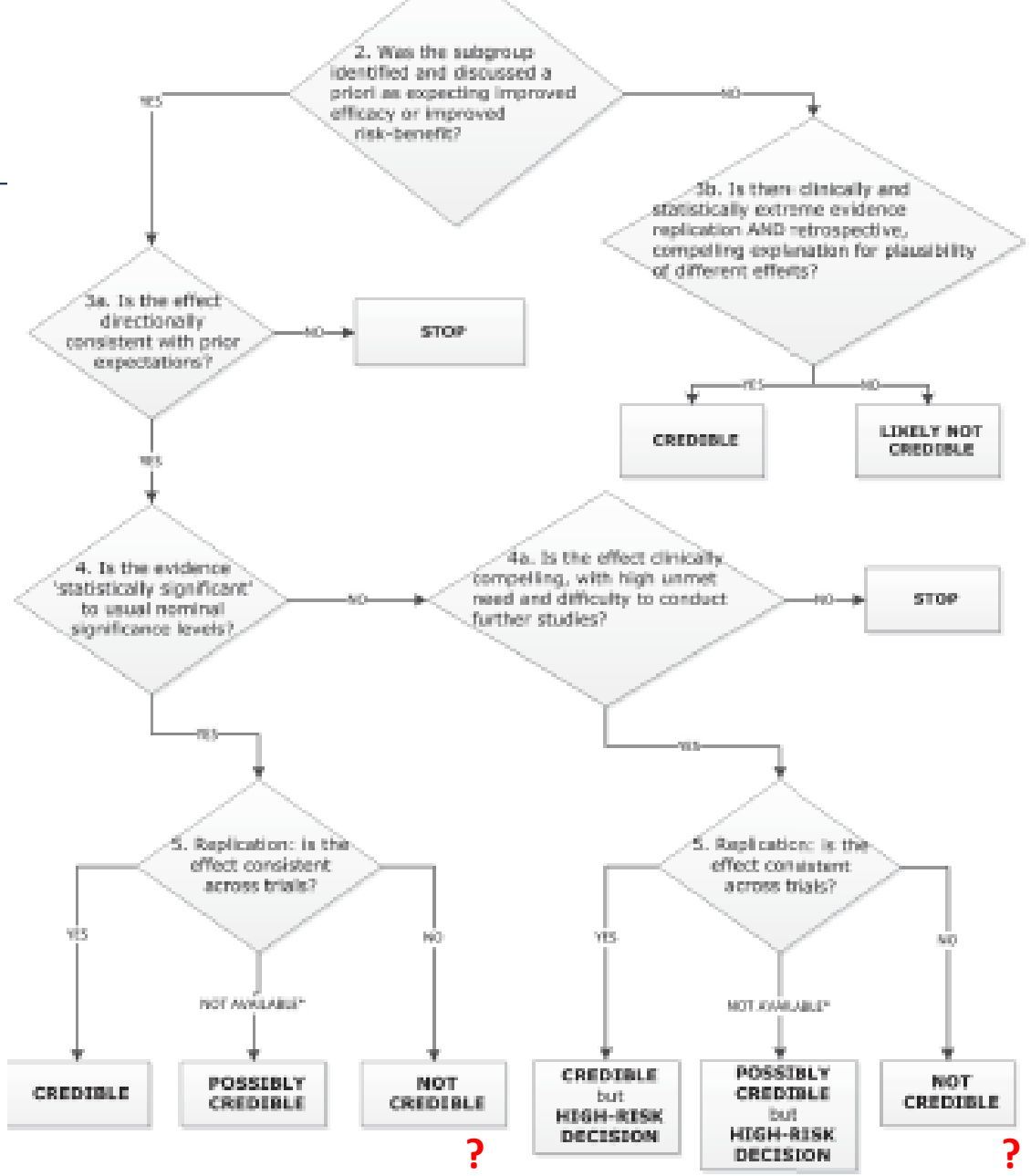
- Proposed approach: for a 2-arm trial, perform intensive stratification at the level of 4 subjects per cell with block size 2 or 4, and 2k subjects per cell with block size k or 2k for a k-arm trial; The proper covariate adjustment for the influential factors, if possible, is suggested to increase the efficiency of the analysis

- This means that for a 2-arm study with a total N=128, up to 32 cells for stratification can be used to allow about up to 5 stratification factors each of which has 2 levels. For a large cardiovascular trial with a total N=16,384 subjects, 4,096 cells are allowed to accommodate up to 12 stratification factors each of which has 2 levels.

- Proper adjustment of the identified factors is recommended. In addition , good subgroup analysis practice with pre-specification, multiplicity adjustment, and others should be followed

Conclusion remark

- A flow chart in EMA guideline (2014)
- Emphasis on consistency multiplicity and clinical plausibility
- Question marks for doubt



Conclusion remark

- This presentation touches upon the inherent characteristics of subgroup analysis related to its credibility

- The most difficult issue in subgroup analysis is identified as the potential group incomparability

- Use of intensive stratification to address the issue is proposed

- Intensive stratification with proper covariate adjustment could be an effective way to improve the quality and credibility of subgroup analysis

- Simulations can be conducted to help to determine the best implementation strategy

- The goal is to created more BELIEVABLE and JUSTIFIABLE subgroup analyses BY DESIGN and AT THE DESIGN STAGE

Reference

- Zhang S, Liang F, Li W, and Hu X (2015), Subgroup Analysis in Reporting of Phase III Clinical Trials in Solid Tumors, J. of Clinical Oncology, 33, 1697-1702

- Draft Guidance for Industry Enrichment Design Clinical Trials for Drugs and Biologics, FDA (2012)

- Sleight P (2000), Subgroup Analysis in Clinical Trial: Fun to look at – but Don't Believe them!, Curr Control Trials Cardiovascular Med, 1, 25-27

- Lagakos S (2006), The Challenge of Subgroup Analysis –Reporting without Distorting, NEJM, 354, 1667-1669

- Guideline on the investigation of subgroups in confirmatory clinical trials. EMA (2014)

- ISIS-2 Collaborative Group (1988), Randomized Trial of Intravenous Streptokinase, Oral Aspirin, Both or Neither among 17817 Cases of Suspected Acute Myocardial Infarction: ISIS-2, The Lancet, 332, 349-360

- Rothwell P (2005), Subgroup Analysis in Randomized Controlled Trials: Importance, Indications and Interpretation, The Lancet, 365, 176-186

[

Thank you