European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 07, Issue 06, 2020

Subgroup Analysis: Approaches, Challenges and Solutions in Biopharmaceutical

Bandi Ramanjineyulu

Manager, Clinical Statistics, GlaxoSmithKline, Bangalore, Karnataka, India

Email id : ramanji.bandi@gmail.com

Abstract:

Subgroup analysis is designed to evaluate whether an intervention has differing effects according to baseline characteristics of participants in clinical trials. This approach can help identify subgroups of patient populations, within a single trial or across multiple trials (meta-analysis), that may benefit from the intervention or can be hypothesis-generating for future trials. Subgroups can also be defined by variables that are prognostic of clinical outcomes or predictive of better treatment effect, such as disease severity, previous therapies, genotype, and biomarker status. Using the same definition of subgroup enables comparison of outcomes between similar subgroups across different clinical trials. This review will provide why we need subgroup analysis, challenges and possible solutions and pitfalls to avoid for investigators involved in the design, conduct, or interpretation of subgroup analyses.

Keywords

Metal-analysis, genotype, biomarker

Introduction:

In clinical research, the safety and efficacy of an experimental treatment is usually assessed by the average treatment effect in the entire patient population. However, in large clinical trials, actual safety and efficacy might vary across patient subpopulations due to differences in some patient characteristics. Since the 1980's, FDA has encouraged and discussed subgroup analyses in various guidance documents. As a result, in the past decades, the drug development process has been improved by the prospect of striking a delicate balance between assessing both average treatment effect and individual treatment effect.

Some subgroups are naturally defined (male / female). Ordinal subgroups have a natural ordering, such as disease severity. Categorical subgroups do not have a natural ordering. However sometimes we create categories from a continuous measure (blood eosinophil count). Intrinsic

ISSN 2515-8260 Volume 07, Issue 06, 2020

characteristics are related to the patient. They may define the patient's physiology, disease pathology or characteristics that are related in some manner to the mechanism of the study drug. Extrinsic characteristics are related to the environment in which the patient sits. An example of an important extrinsic subgroup to regulators is geographic region. This is an important extrinsic factor which can reflect different medical practices, and combining countries into regions requires justification. For example, for a submission to the PDMA in Japan using global trials, there needs to be sufficient Japanese patients recruited in order to demonstrate efficacy and safety. Note that only pre-treatment characteristics qualify as candidates for subgroup analysis. Analyses categorising patients based on measurement collected after randomisation, which may have been influenced by the randomised treatment, requires complex statistical methods.

Why do we need Subgroup Analysis:

Subgroup analyses are **unavoidable** and are needed by many of our stakeholders. Subgroup analyses attempt to answer **important questions** about our medicines, and which types of patients will benefit.

Regulators

We start by considering regulatory authorities who review the clinical data and decide whether to approve the drug based on the benefit-risk profile. The types of patients recruited into the clinical trials needs to reflect the target population who will receive the approved medicine. This is defined in the medicine "label" or prescribing information, and a common strategy is to target a broad range of patients within a disease area, to maximize the number of patients who could potentially receive the medicine. This means that we might have quite a lot of diversity or heterogeneity in the population, so regulators require evidence that there is also a positive benefit-to-risk profile in important subgroups of patients. This is where subgroup analyses become important.

An alternative approach might be where a medicine is targeted at a highly specific group of patients with a particular genotype or biomarker. This is sometimes referred to as personalized medicine. Subgroup analyses are important here as well, because the aim is to show efficacy in the specific target group, but also demonstrate an absence of benefit in patients outside this group.

Payers

Achieving regulatory approval does not guarantee that the medicine will be available to patients. We often need to submit clinical data to payers for the medicine to be made available to prescribers through public or private healthcare providers. Payers have a slightly different perspective to regulators and are interested in whether the drug is cost-effective compared to treatments currently available.

In the case of drug x, although the product had received regulatory approval, the reimbursement agency in Europe only agreed to reimburse the product for a subset of patients with severe asthma,

ISSN 2515-8260 Volume 07, Issue 06, 2020

based on a lab test. This decision was based on subgroup analyses that demonstrated a higher level of cost-effectiveness in this subgroup.
Some payers define methods that are lacking in scientific rigor, such as requiring large numbers of subgroup analyses, which creates problems as. However, there's often little flexibility allowed in this case.

Healthcare Professionals
Once a medicine is approved and reimbursed, healthcare professionals require evidence that the medicine will benefit the patients under their care. In addition to the prescribing information, sometimes additional evidence needs to be provided through publications in medical journals and at therapeutic congresses.

In the drug x example, although regulatory approval and reimbursement hurdles had been cleared, a question arose in the medical community of whether the drug is effective in patients with various types of allergen sensitivity. This question had not been addressed in the original trials and required post hoc subgroup analyses.

Different Approaches of Subgroup Analysis:

Each analysis is independent and does not include any information from the other analyses. We should use the same statistical methods (covariates) as the main analysis on the whole population.

Stratified analysis: The original analysis is repeated separately within each subgroup.

Interaction test: A formal statistical test of heterogeneity between the subgroups, producing a p-value.

You may be familiar with hypothesis tests which answer the question "is there an effect of treatment compared to the control group"? However, in this case, we are assuming there is an overall treatment effect, and the hypothesis test is answering a different question: "is this treatment effect observed consistently between different subgroups of patients"?

Like other hypothesis tests, a small p-value allows us to reject the null hypothesis. Unlike the analysis of the treatment effect in the overall population, which normally concludes statistical significance if the p-value is below 0.05, an interaction test sometimes uses a higher threshold, i.e. concludes there is a significant difference if the p-value is below 0.10. Whatever, the threshold should be pre-specified in the analysis plan. When the effect of treatment is not the same for every level of a pre-treatment (baseline) characteristic, we call that a treatment interaction.

Continuous analysis: A continuous variable to be considered without having to define categories.

ISSN 2515-8260 Volume 07, Issue 06, 2020

Challenges and Solutions:

One of the common themes is "seeking the truth". Actually, we can never know the truth, where the "truth" may be, for example whether there is truly an effect a treatment compared to a control group. The "truth" is where or not there is truly a difference in treatment effect between the groups.

In statistics, we take a sample of data and carry out a statistical test by comparing the p-value to a threshold value. This allows us to make inferences about the "truth". However, conclusions from an analysis may or may not be correct.

True Positive

The truth is that there is a difference between the groups, and the interaction test rejected the null hypothesis, leading to a correct conclusion.

True Negative

The truth is that there is no difference between the groups, and the interaction test failed to reject the null hypothesis. This was also a correct conclusion.

False Negatives

The truth is that there is a difference between the groups, but the interaction test failed to reject the null hypothesis.

Usually, clinical trials are powered to detect a treatment effect in the overall population, and do not have sufficient sample size to detect treatment differences between subgroups of patients. Therefore, it is unlikely to detect such differences if they exist, unless differences between subgroups are particularly large. This problem needs to be addressed at the planning stage.

Solution 1. Firstly, we could increase the threshold value against which the p-value is compared. Discuss this issue with your statistician and consider that fact that this will also increase the rate of false positives, so would need to be made in conjunction with other methods.

Solution 2. We could increase the power by conducting a meta-analysis combining data from more than one study. This is a possible approach if the trials were consistent in terms of patient population, inclusion criteria, the endpoint of interest and study design.

Solution 3. We might be interested in comparing a treatment difference in the overall population with the effect in patients in a targeted subgroup. For example, the subgroup may be patients with a positive result in a diagnostic test or biomarker. This situation sometimes occurs particularly in oncology trials, and in this situation, the study would be powered to test for a differential effect in the subgroup.

False Positives

The truth is that there is no difference between the groups, but the interaction test rejected the null hypothesis.

ISSN 2515-8260 Volume 07, Issue 06, 2020

An importantly characteristic of this scenario is that the more tests we perform, the higher the probability of a false positive. This is known as multiplicity. The more subgroups that are included, the more often we will encounter spurious results due to chance.

Solution: This problem is also addressed at the planning stage. Study teams should first consider important baseline characteristics. These are factors expected to modify any treatment effect, based on mechanism of action, other scientific rationale and experience from previous trials. This should include a relatively small number of key subgroups, specified in the analysis plan, including relevant information (direction of predicted effects, references)

By pre-specifying important subgroups, and documenting the rationale for expected effect, this gives a level of credibility to our conclusions if we observe a treatment interaction in the study results.

Confounding

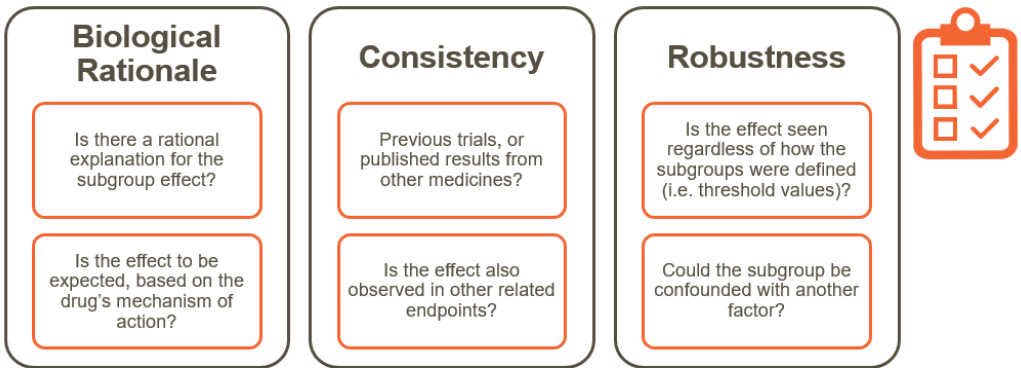
If there's a factor (other than the subgroup characteristic) which impacts the treatment effect, it is known as a confounding factor.

Example if age was suspected to be a confounding variable, i.e., treatment effect may be affected by age. Because there's an imbalance in age between subgroups, the results are difficult to interpret. Also, there may be an imbalance in the confounding variable between treatment groups within a subgroup, leading to a biased estimate.

Solution: The confounder should be adjusted for in the analysis. Where there are baseline characteristics that may be correlated, an alternative approach is exploratory modelling.

Credibility

If differences between subgroups are observed, consider the credibility of the results:



Once the results are obtained, we need to draw conclusions from any quantitative or qualitative interactions that have been observed. The credibility of results from subgroup analyses can be

ISSN 2515-8260 Volume 07, Issue 06, 2020

questioned, both in terms of internal study reporting and also publications in the medical literature. This is especially true of post hoc subgroup analyses.

The credibility is increased if a **Biological Rationale** has already been defined and pre-specified before the study results are produced.

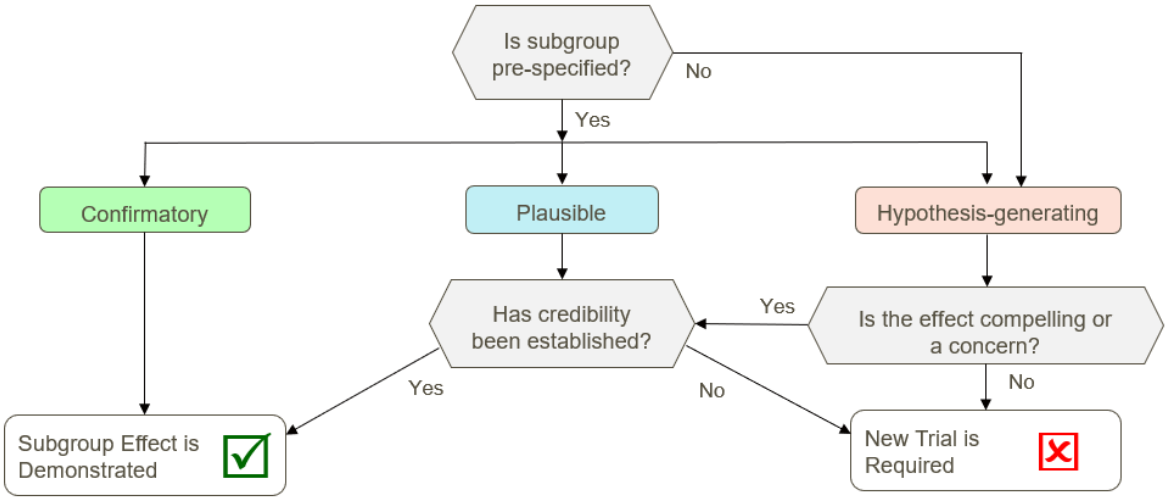
If a subgroup effect is real, it would be expected to show **Consistency** between trials in the same development programme, or other drugs in the same class with the same mechanism of action. Primary and secondary endpoints are often correlated, so an interaction in one endpoint would be expected to be observed in other endpoints, even if the effect is in the same direction but without a significant interaction.

Continuous variables such as age need to be categorised based on threshold values, and these needs to be justified. To determine the **robustness** of the results, we may need to repeat the analysis using a different threshold, to ensure that we the overall conclusions are consistent, regardless of the cut-off point, or to define at which cut-point the conclusions change. We also need to consider the possibility of confounding.

In general, when interpreting the results of subgroup analyses, for example in a Clinical Study Report or article in a medical journal, it's important to avoid overinterpretation of any subgroup effects. I think we've all seen examples of people creating a convincing narrative to fit unexpected results.

Discussion of subgroup effects should include cautious wording, and acknowledge the limitations of the analysis. Reference should be made to other studies that might provide external validity.

Decision Tree:



ISSN 2515-8260 Volume 07, Issue 06, 2020

Confirmatory

This is where the study is designed and powered to test hypotheses about specific subgroups. An example might be an oncology study, where the objective is to confirm that patients testing positive for a genetic marker have a higher response compared to negative patients. By adequately powering the study to compare subgroups, this reduces the false negative rate.

Plausibility

This is a more common situation where the study has been powered to detect a treatment effect in the overall population. Therefore, the study is underpowered for an interaction test, but may show a significant interaction if the difference between groups is large. A small number of key subgroups is defined where there is a plausible rationale for observing a differential treatment effect between subgroups, based on the mechanism of action or experience from previous studies.

Hypothesis-generating

These are the remaining subgroups, where there is no a priori expectation of an interaction. This includes subgroup analyses mandated by regulators. For example, the FDA requires subgroup analyses for age, ethnicity, race, and sex, while the G-BA (Germany HTA agency) expects subgroup analyses for age, gender, country, disease severity for all "patient-relevant" endpoints. Hypothesis-generating subgroup analysis are also relevant to early phase studies, e.g. study of biomarkers. Consider sensitivity analyses using different threshold values.

Conclusion:

We've also seen that there are a number of challenges with subgroup analyses, and we've provided some solutions. Described about the importance of doing an interaction test to compare treatment effects between subgroups, and the possibility of False Positives, leading to spurious results. A solution is to focus on a small number of key subgroups and pre-specify these at the planning stage.

Subgroup analyses are often underpowered, so the interaction test may not show a difference between the groups, even if one exists. To address this problem, we can increase the p-value threshold used for the interaction test, but discuss this with your statistician as this may also increase the false positive rate. In some situations, the study may be designed specifically to address effectiveness in an important subgroup, in which case the study should be designed to answer this question and adequately powered.

Results of subgroup analyses can often lack credibility, especially where subgroups have been defined post hoc. subgroups should be pre-specified where possible, with a scientific rationale documented in the protocol or analysis plan. Results also gain credibility if they are reproducible between a number of studies.

ISSN 2515-8260 Volume 07, Issue 06, 2020

Confounding can be addressed by investigating whether other variables are associated with subgroup factors, and whether it is these rather than the nominal subgroup which may be driving the effect. Discuss with your statistician whether addition modelling work may provide further insights.

If you take away only one key message from this paper, it's that clinical teams should be thinking about subgroup analyses as early as possible, ideally when developing the protocol. Think about key subgroups that are expected to show a treatment interaction, and document as much detail as possible prior to conducting the study. Involve your project statistician as early as possible.

References:

1. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. Stat Med. 1999;18:1903–1942.

2. Food and Drug Administration. Enrichment strategies for clinical trials support approval of human drugs and biological products. https://www.fda.gov/ regulatory-information/search-fda-guidance-documents/ enrichment-strategies-clinical-trials-support-approvalhuman-drugs-and-biological-products. Accessed January 15, 2021.

3. European Medicines Agency. Guideline on the investigation of subgroups in confirmatory clinical trials. https:// www.ema.europa.eu/en/documents/scientific-guideline/ guideline-investigation-subgroups-confirmatory-clinicaltrials_en.pdf. Accessed January 15, 2021.

4. Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. J Biopharm Stat. 2016;26:71–98.

5. Alosh M, Huque MF, Bretz F, D'Agostino RB Sr. Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials. Stat Med. 2017;36:1334– 1360.

6. Lipkovich I, Dmitrienko A, D'Agostino RB Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med. 2017;36:36–196.

7. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332:1080.

8. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009;361:947–957.

9. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351:2817–2826.