On the Utility of Subgroup Analyses in Confirmatory Clinical Trials

EMA Expert Workshop on Subgroup Analyses

7-Nov-2014

Brian A. Millen, Ph.D.

Lilly

Answers That Matter.

Outline

Classification of Subgroup Analyses from Confirmatory Clinical Trials

Proposal for Scenario 1

- General Method
- Implementation Considerations
- Example

Proposals for Scenarios 2 and 3

Summary Comments

Classifying Subgroup Analyses

Confirmatory Subgroup Analyses

- Involve well-defined subpopulations and pre-defined analyses yielding valid inference on the subpopulation(s)
- These analyses make up the primary or key (“gated”) secondary analyses (objectives) of the trial
- Strong control of fwer

Exploratory Subgroup Analyses

- Non-confirmatory analyses
- Analyses separate from the primary or key secondary objectives of the trial

1: supportive analyses

- Offered to support the primary inference
- Based on some *a priori* hypothesis of subpopulation effects

2: discovery analyses

- Used to find potentially viable subpopulations (in a data-driven manner)

Classifying Subgroup Analyses

Confirmatory Subgroup Analyses

- Involve well-defined subpopulations and pre-defined analyses to allow inference on the subpopulation(s)
- These analyses make up the primary or key (“gated”) secondary analyses (objectives) of the trial
- Strong control of fwer

Exploratory Subgroup Analyses

- Analyses separate from the primary or key secondary objectives of the trial

1: supportive analyses

- Offered to support the primary inference
- Based on some *a priori* hypothesis of subpopulation effects

2: discovery analyses

- Used to find potentially viable subpopulations (in a data-driven manner)

Context determines which of these exploratory subgroup analyses is applicable

Setting: Scenario 1

Quote from Draft Guideline:

> The clinical data presented are overall statistically persuasive with therapeutic efficacy demonstrated globally. It is of interest to verify that the conclusions of therapeutic efficacy (and safety) apply consistently across subgroups of the clinical trial population

Setting: Scenario 1

Quote from Draft Guideline:

> The clinical data presented are overall statistically persuasive with therapeutic efficacy demonstrated globally. It is of interest to verify that the conclusions of therapeutic efficacy (and safety) apply consistently across subgroups of the clinical trial population

TENSION

Assume homogeneity unless there is significant evidence otherwise

Assume heterogeneity unless evidence supports homogeneity

How may we address consistency across subgroups?

- Goal: provide helpful information for prescribers and patients
  - Requires …
    - Credibility of the conclusions
    - Must minimize important errors
    - some rigor in the process

One Proposal

Influence Condition

- Introduced for use in confirmatory multipopulation tailoring trials (Millen et al, 2014a,b, 2012)
- May have applicability for the Scenario 1 discussion

The Principle

- Let O = G+ U G-
- Then, given primary inference of beneficial effect (efficacy) in population O, to support a broad indication for the treatment in population O,
- The beneficial effect must not be limited to only the G+ subpopulation

Important Notes

- Requires *a priori* hypothesis for a marker (subgroup and its complement)
- There is not a requirement of equivalent effects.
- The requirement is for each subpop (G+ and G-) to have a positive effect.

Influence Condition

Application of the Influence Condition

- Looking for evidence of positive effect in the individual subgroups
- Millen et al (2014a,b) propose metrics for this
- Bayesian posterior probability
  - $Pr(\theta_{-} > \lambda \mid data) > \gamma$
  - θ_{-} : treatment comparison parameter for marker negative subpop or "least benefitted" subpop
  - λ : benefit threshold (0, 1, other?)
  - γ : evidence threshold

Influence Condition

Application Details

- Setting the parameters
  - Control of influence error rates (Rothman et al 2012; Millen et al 2014a,b)
    - To what level?
- For what subpopulation?
  - There should exist a reasonable prior hypothesis of
  - Size of subpopulation relevant?
- Choice of prior (non-informative vs. informative)
  - informative vs. non-informative
    - based on earlier trials of the drug, external/literature data)?
- Impact on Trial Design
  - Trial is now sized to meet the multiple objectives of overall effect and influence condition evaluation
    - Feasible?
    - At the trial level? Or at the program level?

Motivating Example

Consider a clinical trial with the following assumptions or details

- 2 treatment arms: Drug vs Control
- Primary Endpoint is the difference of treatment means, θ.
- There is a hypothesis that patients in G+ may be better responders to drug than patients in G-
- Apply influence condition as below. Satisfied if

$$\Pr(\theta_{-} > 0 \mid \text{data}) \geq 0.75$$

Example

Operating characteristics

Overall sample size per arm	Relative size of subpop (G+)	θ+	θ-	Influence condition satisfied
133	60%	0.4	0.4	91.6%
200				97.0%
133	85%	0.4	0.4	71.5%
200				80.5%

Example

Operating characteristics

Overall sample size per arm	Relative size of subpop (G+)	θ+	θ-	Influence condition satisfied
133	85%	0.4	0.4	71.5%
200				80.5%
133	85%	0.418	0.3	58.9%
200				68.4%

Application Considerations

Evidence threshold

- 3:1, 2:1 generally?

Size of subpops

- Feasible for small subpops?
- Oversampling/enrichment to increase amount of subpop data available?
  - Complicates reporting overall pop results

Control of error rates

- False Positive (Influence errors) and False Negative

Impact on design / feasibility

Other General Considerations

Important that Scenario 1 evaluations are conducted for very **few** (e.g., 1 or 2) potential markers

- False positive concerns with multiplicity
- Impact on design, feasibility
- Should be done where there is prior hypothesis of potentially significant heterogeneity

Sponsor-regulatory alignment

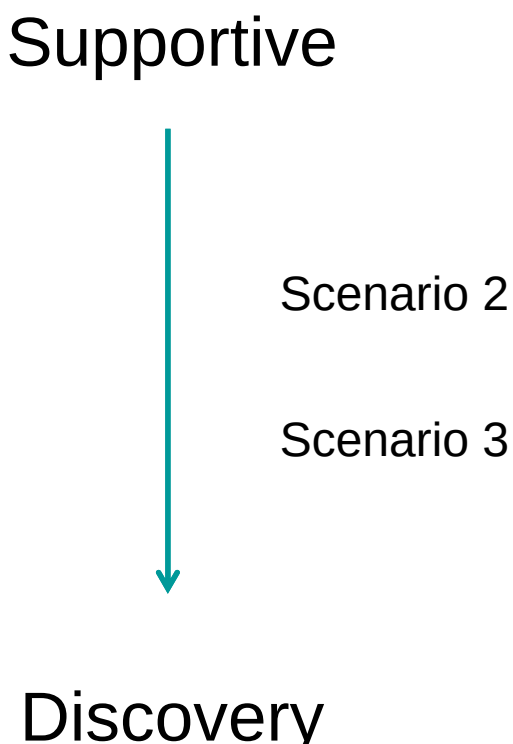
- Pre-defined decision criteria are needed
- Not feasible to have SAWP meeting for every development program. Thus, detailed general guidance will be needed as soon as is practical

Hypotheses of differential effects should be discussed in SAP, rather than in protocol

- Potential to bias investigators when using protocol
- Ability to be flexible and learn (from external sources) while trial is underway

ansition

ansition



tting: Scenario 3

ntext: Overall population result is negative, but there s desire to find subpopulation(s) for which there is a ositive B-R (and potential regulatory action)

posals

irect assessment of predefined subpop (if one existed)

ubgroup ID approaches

- Methodology fully predefined and automated (not requiring human intervention/judgment)
- Lilly approach: appendix of SAP
  - Machine learning tools
  - ‘honest’ estimates

tting: Scenario 2

text: Overall population result is positive, but there is esire to find subpopulation(s) for which there is an nproved B-R (and potential regulatory action)

posals

irect assessment of predefined subpop (if one existed)

ssess Subpops of more severe patients (risk)

- Using natural medical/clinical definitions
- Finding a cut-point along a biomarker
- Credibility of the resulting subpopulation is straight-forward

ubgroup ID approaches

- Methodology fully predefined and automated (not requiring human intervention/judgment)
- Lilly approach: appendix of SAP
  - Machine learning tools

osing Comments

nplex issues

ence trial sponsors and [regulators] are put in a difficult position: whether to ccept an assumption of homogeneity and disregard … plausible findings in ubgroups, or whether to anticipate some heterogeneity and, with appropriate aution and investigation, attempt to use the results of subgroup analyses as one iece of evidence to inform decision making." (129- 133)

ortant to **recognize risks** of subgroup analyses and ppropriately **limit use**

ote that some sections seem to not adhere to this idea

onsideration of feasibility, analyses at trial level may not be formative (particularly for **small subsets**)

earch is needed

he proposals offered here require further research to increase understanding of perating characteristics and develop instructive guidance

esearch into methods in support of the aims of the guidance. The draft

eferences

en, B. A., Dmitrienko, A., Ruberg, S., Shen, L. (2012). A tatistical framework for decision making in confirmatory ultipopulation tailoring clinical trials. *Drug Information Journal* 6:647–656.

en, B. A., Dmitrienko, A., Mandrekar, S. J., Zhang, Z., Villiams, D. (2014).Multipopulation tailoring clinical trials: esign, analysis, and inference considerations.*Therapeutic nnovation and Regulatory Science*.

en, B. A., Dmitrienko, A, Song, G. (2014) Bayesian ssessment of the Influence and Interaction Conditions in Multipopulation Tailoring Cinical Trials. *Journal of iopharmaceutical Statistics* 24: 94-109.

hmann, M. D., Zhang, J. J., Lu, L., Fleming, T. R. (2012).

ACK-UP SLIDES