BfArM
Federal Institute for Drugs and Medical Devices

The update of the multiplicity guideline

Norbert Benda
Federal Institute for Drugs and Medical Devices (BfArM), Bonn

Disclaimer:
Views expressed in this presentation are the author's personal views and not necessarily the views of BfArM

EMA Points to Consider on Multiplicity Issues

- Adopted 2002
- Describes
  - when adjustment is needed, when is it not
  - claims from multiple secondary endpoints
  - conclusions from subgroup analyses
  - interpretation of responder analyses in addition to analysis of original endpoint
  - handling of composite endpoints

EMA PtC: When adjustment is (not) needed

not needed, e.g.:

- Two or more co-primary endpoints
  - Significance needed for all endpoints
- Two or more primary endpoints ranked according to clinical relevance
- Multiple analysis sets

needed:

- Multiple ways of study success

→ *Terminology ("adjustment") to be clarified*

EMA PtC: Subgroups

- Conclusions for subgroups require
  - prespecification and appropriate multiple test procedures
- Claim for subgroup unlikely to be accepted
  - if overall populations fails to show significant effect
- Heterogeneous results among subgroups
  - may lead to the restriction to certain subgroups

→ *New guideline on subgroup analysis in preparation*

EMA PtC: Secondary endpoints

- No claims intended:
  - No adjustment, explorative interpretation
- Claims intended
  - Confirmatory testing
  - Multiple testing procedures needed
  - Only to be tested if primary hypotheses are rejected

→ *Multiple primary endpoints need further specification on how to step from primary to secondary*

EMA PtC:
In general

- Basic principles still valid
  - Type 1 error control related to study success
- Clarifications needed
  - role of additional claims
  - terminology
- Additional issues to be added
  - confidence intervals
  - more complex multiplicity frameworks

Concept paper on the need for a guideline on multiplicity

Proposed topics to be discussed

- Combinations of different sources of multiplicity
- Usefulness and limitations of new strategies
- Multiplicity issues in confirmatory conclusions in subgroups
- Multiplicity issues arising from interim decisions
- Multiplicity in multiregional developments
- Simultaneous confidence intervals corresponding to multiple test procedures

Basic principles in drug approvals

1. Demonstrate efficacy (study success)
2. Show favourable benefit risk
3. Additional claims need to be demonstrated in a confirmatory way after general efficacy (1) has been shown (*PtC on Multiplicity*)

Demonstrating efficacy

- Define win situation, e.g
  - A one primary endpoint/hypotheses
  - A and B co-primary endpoints
  - A or B multiple possibilities
- Prob(win | no effect) ≤ α
  - Prob(drug approved | drug ineffective) ≤ α
  - "Specificity":
    Prob(drug not approved | drug ineffective) ≥ 1-α

Weak vs strong FWER

Familywise error rate (FWER):

- Weak:
  - Prob(drug approved | no effect in none of the questions)
- Strong:
  - Prob(drug approved | no effect in any combination of the questions)

Multiplicity due to

- multiple endpoints
- multiple populations / subgroups
- multiple doses, regimens
- multiple looks
- etc.

and

- combinations of all this

Secondary endpoints

- additional claims only relevant if efficacy shown in primary endpoint(s)
  - parallel/ improved gatekeeping useful
- role of secondary endpoints to be clarified
  - when is confirmation needed in secondary endpoints?
  - confirmed claim vs descriptive label

Counterintuitive results

- apparently highly effective endpoints may not be claimed due to low weight / "backmost position in the queue"
- lack of frequentist thinking in the assessment
  - simplified situation:
    - "False hierarchy":
      - Testing low dose first, then high dose
      - $p_{low} = 0.2$, $p_{high} = 0.0001$
  - Trial failed
  - but clinical assessment often ignores design:
    ("... *you clearly see that the high dose is effective ...* ")

Benefit risk profile

- Additional demonstrated claims improve benefit risk profile:
  - Complex hypothesis framework may lead to a better assessment by answering more questions
  - However:
    - Multiplicity adjustment could penalize complex frameworks compared to simple ones
    - Conclusions may depend on multiplicity procedure applied
  - Impact of secondary endpoints on benefit risk relevant to the choice of multiplicity procedure
- Evaluation of benefit risk profile asks for proper confidence intervals

Issues to be resolved

- Role of secondary endpoints to be clarified
- Transparency needed
  - not only results themselves decide on success but also ways to get there
- Reasonable confidence intervals
  - Stepwise procedures do not allow for simple and informative simultaneous confidence intervals
  - Bonferroni-like methods seem to be the only remedy
- Relation to benefit risk assessment

Comments on the Concept paper

Comments received from

- EFPIA
- QSPI multiplicity working group
  - industry-wide working group sponsored by the industry group “Quantitative Sciences in the Pharmaceutical Industry
- Ohio State University
- Lancaster University

Major comments on

- Scope of the guideline to be defined
  - early phase ?, dose finding ?, etc.
- Type of error control
  - other concepts than FWER , false discovery rate ?
  - expected # false claims ?
- Simultaneous confidence intervals
  - use of partitioning principle
  - no informative ci for “all powerful multiple test procedures“, only for Bonferroni-like tests

Major comments on

- Relationship between hypotheses
  - use of clinical information to exclude interpretational problems
- Role of secondary endpoints
  - conditions for labelling claims to be defined
  - when do we have to consider secondary endpoints in the FWER ?
- Subgroups
  - need for predefined subgroups ?
  - significant effect in the overall needed ?
  - conditions for a restrictions on subgroups

Major comments on

- Objectives to be clarified
  - primary objective for trial success
  - secondary for labelling claims
  - descriptive / exploratory
  - differentiate
    - trial successful if at least one test significant
    - trial successful if all tests are significant
    - trial successful if global test is significant
- Terminology
  - hierarchical testing, co-primary endpoints: special multiple testing strategy instead of “no adjustment needed“
  - multiplicity more than “adjustment“

Other comments on

- Multiplicity in safety assessments
- Composite endpoints
- Interim decisions
  - adjustment for secondary endpoints at interim
- Multiplicity in parallel studies
- Multiple doses
  - conclusion for individual dose needed
- Multiregional trials
  - different SAPs for different agencies
  - consistency between regions / effect in EU population