Federal Institute
for Drugs
and Medical Devices

Regulatory Issues with Multiplicity in Drug Approval and Current Controversies

Norbert Benda

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

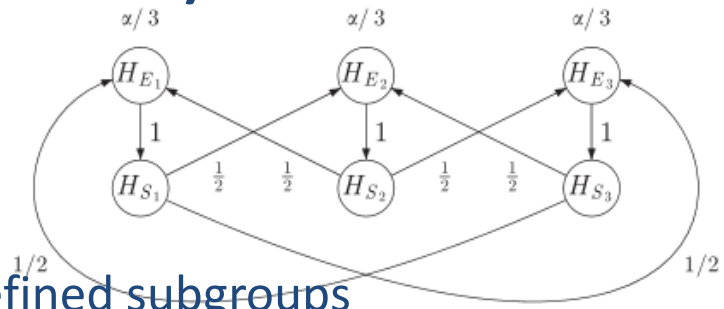
Overview

- Multiplicity and drug approval
- Success criterion and multiple conclusions
- Main changes
- Selected issues and comments
- Conclusion

History

- Points to Consider (PtC) on Multiplicity (2002)
- Concept paper on the need for a guideline on multiplicity issues in clinical trials (2012)
- Comments from academia and industry
- EMA Workshop on Multiplicity (Nov 2012)
- New guideline on Multiplicity issued 2016

Multiplicity issues and trends (scientific advice procedures + submissions)



- **New methods developed since 2002**
  - e.g. graphical methods, etc.
- **Increasing complexity**
  - increasing investigation of biomarker defined subgroups
  - increased use of complex combinations, e.g.
    - multiple endpoints + multiple dose groups + multiple populations + multiple analyses / interim decisions
  - possibility to investigate multiple options in a single study to increase the chance of a successful study
- **Traditional dose finding concept challenged**
- **Discussions on the need for a confirmatory proof in secondary endpoints**
- **Selection bias in estimation**
- **Use of simultaneous confidence intervals**

Drug approval

- **Approval requirements**
  - Two pivotal Phase III trials
  - Confirmatory proof of efficacy in RCTs
    - selected primary endpoints to describe efficacy
    - control of false positive decisions w.r.t. study success
      - pre-specification of study success by primary endpoints
      - corresponding multiplicity adjustment
  - Descriptive assessment of safety/side effects
    - "sufficient" size of safety database
  - Benefit risk (BR) should be considered positive
    - balancing benefit and risks/efficacy and side effects

Drug approval based on

- **Confirmatory proof of efficacy**
  - Frequentist approach to limit false positive decisions
    - Statistical significance test on primary efficacy endpoints
    - Limit proportions of successful studies by (usually) 2.5% in absence of any effect
    - Different options to define study success
      - e.g. by two co-primary endpoints: both must be significant
    - Control properties (= long-run probabilities) of the decision procedure

Drug approval

- **Confirmatory proof of efficacy**
  - gatekeeping function followed by further assessment
  - combined with a descriptive assessment of safety / benefit risk (BR)

    means
    - positive BR : “best guess” – best available evidence
      - confirmatory proof in further endpoints not required for study success but would strengthen BR assessment
      - structured BR approaches possible but not confirmatory

Success criterion

- **Confirmatory proof of efficacy**
  - Frequentist approach to limit false positive decisions
    - Statistical significance test on primary efficacy endpoints
    - Limit proportions of successful studies in absence of any effect
    - Different options to define study success
      - e.g. by two co-primary endpoints: both must be significant ("and")
      - or two alternative options ("or")
    - Control properties (= long-run probabilities) of the decision procedure
      - frequentist approach, difficult to convey to non-statisticians

Success criterion, multiple conclusions and additional confirmatory statements

- Success criterion:
  - based on primary efficacy endpoints
  - defines study success
  - type-1 error rate limits false decisions of success
- Multiple conclusions
  - different conclusions to be taken for approval
    - e.g. relating to posology and indication
      - efficacy in different dose groups
      - efficacy in different subpopulations
- Additional confirmatory statements
  - may not be needed for primary trial success
  - but may inform BR assessment and labelling

Success criterion

- Proper definition of study success needed
  - define win criterion e.g. by
    - A and B co-primary endpoints

      or
    - A or B multiple possibilities
- Apply study-wise error control in the strong sense to all confirmatory conclusions
  - Prob(false positive conclusion l any combination of null hypotheses) ≤ *a*
- Discussions on “or” criteria, e.g.
  - Study success defined by a significant difference in
    - either progression free survival (PFS) or overall survival (OS)
    - Adjustment
      - e.g. PFS: $\alpha = 0.01$, OS: $\alpha = 0.04$
    - frequently combined with interim analyses

Success criterion and multiple conclusions

- Study success defined by
  - "and" combinations
    - co-primary endpoints/hypotheses
    - rejection of $H_1 \cup H_2$
    - conclusion on A and B
    - single conclusion
  - "or" combinations (endpoints, doses, subpopulations)
    - rejection of $H_1 \cap H_2$
      - study success criterion fulfilled
    - conclusion on A or B
    - but may imply multiple conclusions
      - approval of dose 1 or 2, population 1 or 2
      - single test of $H_1 \cap H_2$ insufficient (e.g. by a pooled analysis)

Main changes from Points to Consider to the new guideline

- **wording**
  - e.g. “study-wise” error control, type-1 error in the strong sense, multiplicity in safety, etc. …
- **safety**
  - in multiple tests as a flagging device to signal a potential risk an adjustment for multiplicity is counterproductive
  - different nature of conclusions
    - regulatory aim to show absence of harm
    - absence of a proof ≠ proof of absence
  - not part of multiplicity considerations covered by the guideline
- **emphasis on study-wise error control regarding**
  - all *confirmatory* conclusions

Main changes from Points to Consider to the new guideline

- **dose finding**
  - changed paradigm
    - from the multiple comparisons to placebo in Phase II studies
      (lowest significant dose may be too high or too low)
    - to a reasonable regression analysis to conclude on a reasonable dose (range) for Phase III
- **EMA workshop on dose finding (2014)**
  - *"the workshop re-emphasised the importance of rigorous, scientific dose finding (relying on model-based estimation, rather than hypothesis testing via pairwise comparison)"*
- **Guideline changed accordingly**

Main changes from Points to Consider to the new guideline

- **use of secondary endpoint**
  - extensive discussions on the need of a confirmatory proof in secondary endpoints (requiring multiplicity adjustment)
  - formal claims associated with a confirmatory proof in secondary endpoints usually not part of the requirements
  - confirmatory proof in secondary endpoints strengthen the benefit risk assessment

Main changes from Points to Consider to the new guideline

- **alternative analysis methods / “modelling”**
  - issues related with data based model/analysis selection
  - confirmatory conclusion require type-1 error control of whole procedure
    - Conclusions on finally selected models may be too optimistic
  - data driven model change discouraged

Main changes from Points to Consider to the new guideline

- **subgroup evaluation**
  - scope of this guideline:
    - multiplicity adjustments in pre-planned subgroup analyses
    - pre-planned (adaptive) subgroup selection
      - Subgroups, null hypotheses and test procedures to be pre-specified
        - to allow for a proper confirmatory conclusion on a specific claim of a beneficial effect in a particular subgroup
    - appropriate pre-planned multiplicity adjustment is needed for an unambiguous confirmatory conclusion.
- reference made to the new *Guideline on the Investigation of Subgroups in Confirmatory Clinical Trials* re. decision making in general

Main changes from Points to Consider to the new guideline

- **multiplicity in estimation: new section**
  - selection bias
  - simultaneous confidence intervals
- **multiple comparisons may lead bias in estimation**
  - e.g. strategy that chooses the treatment with the largest difference to placebo will lead to an overestimation of the treatment effect.
  - If selection is made not on the basis of the treatment effect it may still be based on an endpoint that is correlated with efficacy
  - selection at an (earlier) interim analysis
    - less bias though less informative
  - methods available to reduce selection bias
    - e.g. shrinkage estimation

Main changes from Points to Consider to the new guideline

- **multiplicity in estimation: simultaneous confidence intervals**
  - Informative confidence regions that correspond to multiplicity procedures may not always be available or difficult to derive
  - If the confidence regions do not correspond to the hypothesis testing procedure different conclusions are possible
    - *e.g.* a confidence interval excluding the null hypothesis combined with a non-significant testing result or *vice versa*
- **use of simultaneous confidence intervals**
  - Correspondent confidence intervals difficult to construct
  - "additional multiplicity" with respect to ci and tests
    - test prevails
      - decision to be based on the hypothesis test
    - reporting issues to be discussed

Selected comments received on further issues

- **Give details on adaptive designs**
  - specific guideline on adaptive designs
- **Inclusion of multiple estimand**
  - specific ICH E9 addendum
  - define primary estimand
- **Sample size considerations**
  - not considered within the scope of the GL
- **Specific considerations of non-inferiority and equivalence**
  - to be considered
- **more clarity on secondary endpoints**
  - to be discussed
- **multiple regional requirements**
  - may be considered

Selected comments received on further issues

- **details on specific methods**
  - not within the scope of the GL
- **inclusion of Bayesian methods**
  - currently not intended
- **multiplicity in estimation / confidence intervals**
  - to be enhanced / extended
  - reporting issues controversial
- **composite endpoints**
  - “seems to not fit in with the objectives of the document”
- **give examples**

Further issues

Studies to obtain conditional approval (CMA) at interim

- **Should the type-1 error rate be controlled for CMA and full MA separately or jointly?**
- **CMA may imply marketing for several years before final analysis**
  - controlling type-1 error rate for *any* authorization, i.e. incorporating both, the comparison(s) used for CMA and that (those) for MA in the multiplicity procedure
- **Example: Treatment of non-alcoholic steatohepatitis (NASH)**
  - co-primary endpoints "Composite of complete resolution of steatohepatitis and no worsening of fibrosis stage" and "Composite of one point improvement in fibrosis stage and no worsening of steatohepatitis" considered as surrogate endpoints for interim evaluation as potential basis for CMA
  - hard clinical endpoint for full MA: composite time-to-event endpoint with events death, liver transplantation, cirrhosis-related clinical events, and histological diagnosis of cirrhosis
  - discussed options:
    - splitting of the alpha level between surrogate and clinical endpoint with the possibility for a confirmatory analysis of the hard clinical endpoint and obtaining a full MA even after failing to meet the surrogate endpoint
    - hierarchical testing which allows continuing the study after interim only if the surrogate endpoint is met and obtaining CMA might be possible

Further issues

Multiple SAPs for different authorities

- **Different study success definitions used for FDA and for EMA**
  - different endpoints
  - co-primary vs single primary endpoints
- **Agreement on**
  - separate consideration of each SAP
- **Increased discussions EMA - FDA on consistent requirements**
  - parallel scientific advice procedures with EMA and FDA