www.pei.de

(Draft) Guideline on multiplicity issues in clinical trials

Dr. Benjamin Hofner
Section Biostatistics

Forum Biomedizinische Arzneimittel
14.06.2017

Statistical Testing
Type I and Type II Error

- Statistical tests form the basis of confirmatory studies
- Test if null hypothesis
  H_0: no difference between treatment and placebo
  can be rejected (and thus alternative accepted).
- Tests are construced to minimize wrong decisions

	H_0 true	**H_1 true**
H_0 rejected	**Type I error (α)**	✓
H_0 not rejected	✓	**Type II error (β)**

- Type I Error: P(reject H_0 if H_0 is true) = α
  - usually α set to 5% (two-sided) or 2.5% (one-sided)
- Type II Error: P(not reject H_0 if H_1 is true) = β
  - β minimised
  - 1 - β = Power

Dr. Benjamin Hofner
Section Biostatistics

General problem

- The main scope of the GL is to provide guidance on the confirmatory conclusions which are usually based on the results from pivotal Phase III trials and, to a lesser extent, on Phase II studies. The guideline mainly discusses issues in decision making for a **formal proof of efficacy**.

- It is well known that the likelihood of a **positive chance finding increases** with the number of questions posed, if no actions are taken to protect against the inflation of false positive findings from multiple statistical tests.

➢ **Pre-specification is key:**
Multiplicity considerations must be fully detailed in the study protocol or in the statistical analysis plan.

Dr. Benjamin Hofner
Section Biostatistics

Multiplicity

- For example, if statistical tests are performed on five subgroups, independently of each other and each at a significance level of 2.5% (one-sided), the chance of finding at least one false positive statistically significant test increases to approximately 12%.

  $1 - (1 - 0.025)^5 \approx 0.119$

- Opportunity to choose the most favourable result:
  - Increased risk of false positive effect and hence wrong MA

- Many methods for multiplicity control exist (e.g. Bonferroni)

  Use $\alpha/5 = 0.005 \Rightarrow$ Overall error with corrected α: ≈ 0.0247

- Choice of method should be driven by interpretability of results (and availability of proper confidence intervals)

Dr. Benjamin Hofner
Section Biostatistics

History of the GL

- New guideline strongly builds on previous PtC (from 2002)
- EMA Concept paper on the need for a guideline published in 2012
- Draft GL (2016/17):

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

15 December 2016
EMA/CHMP/44762/2017
Committee for Human Medicinal Products (CHMP)

Guideline on multiplicity issues in clinical trials

Draft

Draft agreed by Biostatistics Working Party (BSWP)	November 2016
Adopted by CHMP for release for consultation	15 December 2016
Start of public consultation	01 April 2017
End of consultation (deadline for comments)	30 June 2017

Major changes

- New sections on
  - multiplicity in estimation, and
  - new approaches in dose finding.

- Furthermore, a lot of smaller changes were made clarifying specific issues and applications of multiplicity corrections.

Dr. Benjamin Hofner
Section Biostatistics

Claims

- The guideline uses the term “claim“:
  - shorthand for a confirmatory conclusion which is then prioritised in trial reporting and used as primary basis for asserting that efficacy or safety has been established

- Claims can only be made
  - if primary endpoint was successful,
  - if secondary endpoints / subgroups were pre-specified, and
  - if they were part of a multiplicity control strategy

- Fun Fact
  - FDA Guideline on multiplicity does not use the term “claim”

Dr. Benjamin Hofner
Section Biostatistics

Main questions answered by GL

Main Sections in GL:

5. Adjustment for multiplicity – when is it necessary and when is it not?
6. How to interpret significance with respect to multiple secondary endpoints and when can a regulatory claim be based on one of these?
7. When can confirmatory conclusions be drawn from a subgroup analysis?
8. How should one interpret the analysis of 'responders' in conjunction with the analysis of raw variables
9. How should composite endpoints be handled statistically with respect to regulatory claims?
10. How should multiplicity issues be addressed in estimation?

Not part of this GL: Interim analyses

- RP on Methodological issues in Confirmatory Clinical Trials planned with an Adaptive Design (CHMP/EWP/2459/02).

5.

Adjustment of elementary hypothesis tests for multiplicity – when is it necessary and when is it not?

Dr. Benjamin Hofner
Section Biostatistics

5.1. Primary Endpoints

Dr. Benjamin Hofner
Section Biostatistics

Single Primary EP

- Only one single primary endpoint (EP; and no key secondary EPs)
  - No control necessary

Dr. Benjamin Hofner
Section Biostatistics

Multiple Primary EPs

- Multiple primary EPs, which **all** need to be significant for study sucess (all other EPs supportive only)
  - ➢ Co-primary endpoints
  - Needed if clinical effect is not defined by single EP
  - No adjustment of significance level needed
  - Power decreased (as more than one EP must be positive)
    - Must be considered at planning stage
  - No intention or opportunity to select the most favourable result

Dr. Benjamin Hofner
Section Biostatistics

Multiple Ranked EPs

- Two or more primary EPs ranked according to clinical relevance
- “One objective is of greatest importance but convincing results in others would clearly add to the value of the treatment”
  - ➢ No adjustment of significance level needed
  - No claim can be based on EPs with lower rank than a test that was not significant
  - Pre-defined ordering avoids any choice in the assessment but must “be pre-specified in the study protocol”
  - Must include a clear specification of the set of hypotheses that need to be significant before the trial is claimed successful
  - (Type II errors are inflated for hypotheses that correspond to endpoints with lower ranks)
- Can be extended to key secondary EPs

Examples (not part of the GL)

Co-primary EPs

- ✓ EP1: p = 0.01
- ✗ EP2: p = 0.06

➢ Trial failed

Pre-defined ranking

- ✓ EP1: p = 0.01
- ✓ EP2: p = 0.04
- ✗ EP3: p = 0.12
- ? EP4: p = 0.001

➢ EP3 and EP4 failed

Dr. Benjamin Hofner
Section Biostatistics

5.2. Multiple Analysis Sets

Dr. Benjamin Hofner
Section Biostatistics

Multiple Analysis Sets

- When the aim is to assess robustness of primary analysis
  - No adjustment needed
  - Primary analysis set must be pre-specified
  - Main purpose of such analyses is to increase confidence in the results obtained from the primary analysis

Dr. Benjamin Hofner
Section Biostatistics

5.3. Multiple Analysis Methods

Dr. Benjamin Hofner
Section Biostatistics

Post hoc choice of analysis model

- Two-step procedure applied with the purpose of selecting a particular statistical technique / test for the main treatment comparison inflates type 1 error
  - This includes model/variable selection approaches
  - Type 1 error of complete strategy needs to be considered
- Any *post hoc* selection of the model is not considered appropriate for a confirmatory Phase III trial
- Blinded selection might also increase the type 1 error
  - It must be “properly justified with respect to type I error control and its potential impact on the treatment effect estimate as regards bias”
- In summary, “such procedures are not recommended. Confirmatory analyses should be fully and precisely pre-defined”

Dr. Benjamin Hofner
Section Biostatistics

5.4. Safety Variables

Dr. Benjamin Hofner
Section Biostatistics

Safety variables

- If part of the confirmatory strategy (for approval or labelling claims), it should not be treated differently from the primary efficacy endpoints
  - see also composite endpoints
- P-values often not useful for AEs/SAEs as effect size, seriousness / severity, etc. need to be taken into account.
- Non-significant safety findings are not a proof of equivalent safety profiles.

- If statistical test procedures are performed to flag / signal a potential risk caused by the investigational drug, adjustment for multiplicity is counterproductive
  - Clinical judgement, PK judgement needed in addition to flagging

5.5. More than two treatment arms

Dr. Benjamin Hofner
Section Biostatistics

The three arm ‘gold standard’ design

- If commonly acknowledged reference drug therapy exists
  - three-arm study with the reference drug, placebo and the investigational drug might be requested.
- (Often) demonstrate
  - superiority of the investigational drug over placebo (proof of efficacy)

  and
  - investigational drug retains, at least, most of the efficacy of the reference drug as compared to placebo (proof of non-inferiority)
- ➢ If study success is defined by both EPs, no formal adjustment needed (and both EPs must be significant)
- ➢ If only efficacy must be shown no adjustment for non-inferiority test needed (if no claim intended; elsewise pre-defined multiplicity adjustment needed)

Dr. Benjamin Hofner
Section Biostatistics

Proof of efficacy for a fixed combination

- Combination therapy
- CPMP guideline (CPMP/EWP/240/95 Rev. 322 1) requires that "each substance of a fixed combination must have documented contribution within the combination"
  - 3 or 4-armed study:

    Monotherapy 1 *vs.* Monotherapy 2 *vs.* Combi (*vs.* Placebo)

    - Superiority of combi over each mono therapy needs to be shown
  - No multiplicity adjustment needed

Dr. Benjamin Hofner
Section Biostatistics

Dose-response studies (new)

- Multiplicity adjustment of the different comparisons (doses) between groups in order to control the study-wise type I error may not be required in a Phase II trial.
  - Valuable achievement: “demonstration of an overall positive correlation of the clinical effect with increasing dose”
  - Estimates and confidence intervals are used for an appropriate interpretation of the dose-response and may be used for the planning of future studies.
- ICH E4 discusses dose-response studies as part of the confirmatory package
    - A pre-specified plan to control the type I error is of importance.
- Phase III studies with multiple doses must control the study-wise type I error

Dr. Benjamin Hofner
Section Biostatistics

6.

How to interpret significance with respect to multiple secondary endpoints and when can a regulatory claim be based on one of these?

Dr. Benjamin Hofner
Section Biostatistics

Secondary endpoints for supportive evidence (expanded)

No claims are intended

- Multiplicity control not mandatory but gives further support
- Ranking might be controversial
  - Common practice to rank endpoints based on the likelihood that the individual null hypothesis can be rejected
  - Ideally the clinical assessment should focus on those endpoints of greater clinical importance
- Even if no formal multiplicity control, specification of few key secondary EPs (out of many secondary EPs) might be beneficial for interpretation and/or the SmPC

Dr. Benjamin Hofner
Section Biostatistics

Secondary endpoints for additional claims

Significant effects in these endpoints can be considered for an additional claim only after the primary objective of the clinical trial has been achieved, and if they were part of the confirmatory strategy.

- Multiplicicity control needed
- Hierarchical testing approach is a common choice
- Other approaches exist
- Same as for mulitple primary EPs

- Depending on the degree of complexity, regulatory dialogue is recommended to assure that the outcome of the procedure can be interpreted in clinical terms.

Dr. Benjamin Hofner
Section Biostatistics

Secondary endpoints indicative of clinical benefit

- EPs for major clinical benefit / important safety issue (*e.g.* mortality) might be considered secondary (e.g. as study is not powered for these EPs)

- If primary EP fails, and effect of this secondary EP much larger than anticipated
  - "further studies would be needed to support the observed beneficial effect".
  - ➢ No confirmation of hypothesis!

- If primary EP successful but major clinical benefit in the wrong direction
  - results might be doubted and a "Marketing Authorisation may not be granted, regardless of whether or not this endpoint was embedded in a confirmatory scheme."

7.

Reliable conclusions from a subgroup analysis, and restriction of the licence to a subgroup

Dr. Benjamin Hofner
Section Biostatistics

Subgroup analysis (shortened)

- Often supportive only
- Claim of a beneficial effect in a particular subgroup requires pre-specification and multiplicity adjustment
- ➢ See CHMP guideline on the Investigation of Subgroups in Confirmatory Clinical Trials (EMA/CHMP/539146/2013).

A licence may be restricted if unexplained strong heterogeneity is found in important sub-populations, or if heterogeneity of the treatment effect can reasonably be assumed but cannot be sufficiently evaluated for important sub-populations.

Dr. Benjamin Hofner
Section Biostatistics

8.

How should one interpret the analysis of ‘responders’ in conjunction with the raw variables?

Supportive Responder Analysis

- Binary (version of) response (e.g. ACRxx)
- Responder/non-responder should be pre-specified in the protocol and should be clinically convincing
- Clinical GLs:
  - Responder analysis should be used in establishing the clinical relevance of the observed effect as an aid to assess efficacy and clinical safety.
  - But loss of information (and hence loss of statistical power)
- If responder analysis is used to allow a judgement on clinical relevance, once a statistically significant treatment effect on the mean level of the primary variable(s) has been established results need not be statistically significant but the difference in the proportions of responders should support a statement that the investigated treatment induces clinically relevant effects.

Dr. Benjamin Hofner
Section Biostatistics

9.

How should composite endpoints be handled statistically with respect to regulatory claims?

Composite endpoint as primary endpoint

- First test overall composite EP
- Additionally analyse single components and clinically relevant groups of components separately (supportive information):
  - Treatment should affect all components similarly
  - Clinically more important components should at least not be affected negatively
- No need for an adjustment for multiplicity provided significance of the primary endpoint is achieved.
- If claims are to be based on (subgroups of) components, this needs to be pre-specified and embedded in a valid confirmatory analysis strategy.

- More details on the construction, analysis and interpretation of composite EPs can be found in the multiplicity GL.

Dr. Benjamin Hofner
Section Biostatistics

10.

Multiplicity issues in estimation (new)

Dr. Benjamin Hofner
Section Biostatistics

Selection bias

- (Usually bias from specific patient or subgroup selection)

- Multiple comparisons may lead to a bias in effect estimate
  - Several treatment groups / doses are compared to placebo
  - Chose treatment with the largest difference to placebo
  - ➢ Overestimation of the corresponding treatment effect (also if based on other measures related to efficacy)

- Selection bias usually lower (but still present) if treatment selection based on interim data (but this data is usually less informative).
  - Selection bias can be reduced via shrinkage estimation or model based analyses.

Confidence intervals

- Confidence intervals (CIs) are another important basis of the clinical interpretation.
  - CIs and tests are interrelated
  - ➢ CIs should also be adjusted for multiplicity
- CIs that correspond to multiplicity procedures may, however, not always be available or may be difficult to derive.
  - Different conclusions are possible
    - CIs excluding the null hypothesis
    - but non-significant testing result or *vice versa*
  - Decide based on test and use conservative CIs such as Bonferroni-corrected intervals

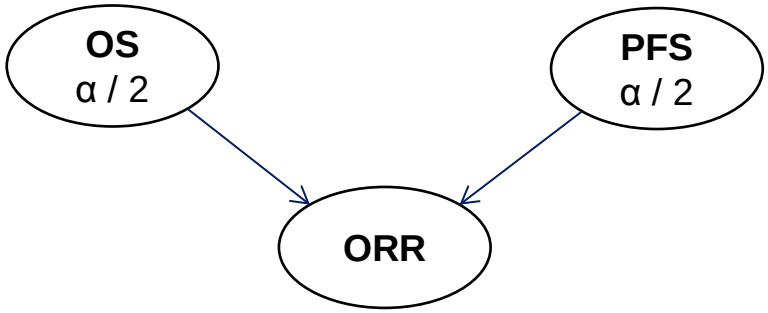
Personal note: CIs are often not corrected in current MAA dossiers

Outlook
(not part of the GL)

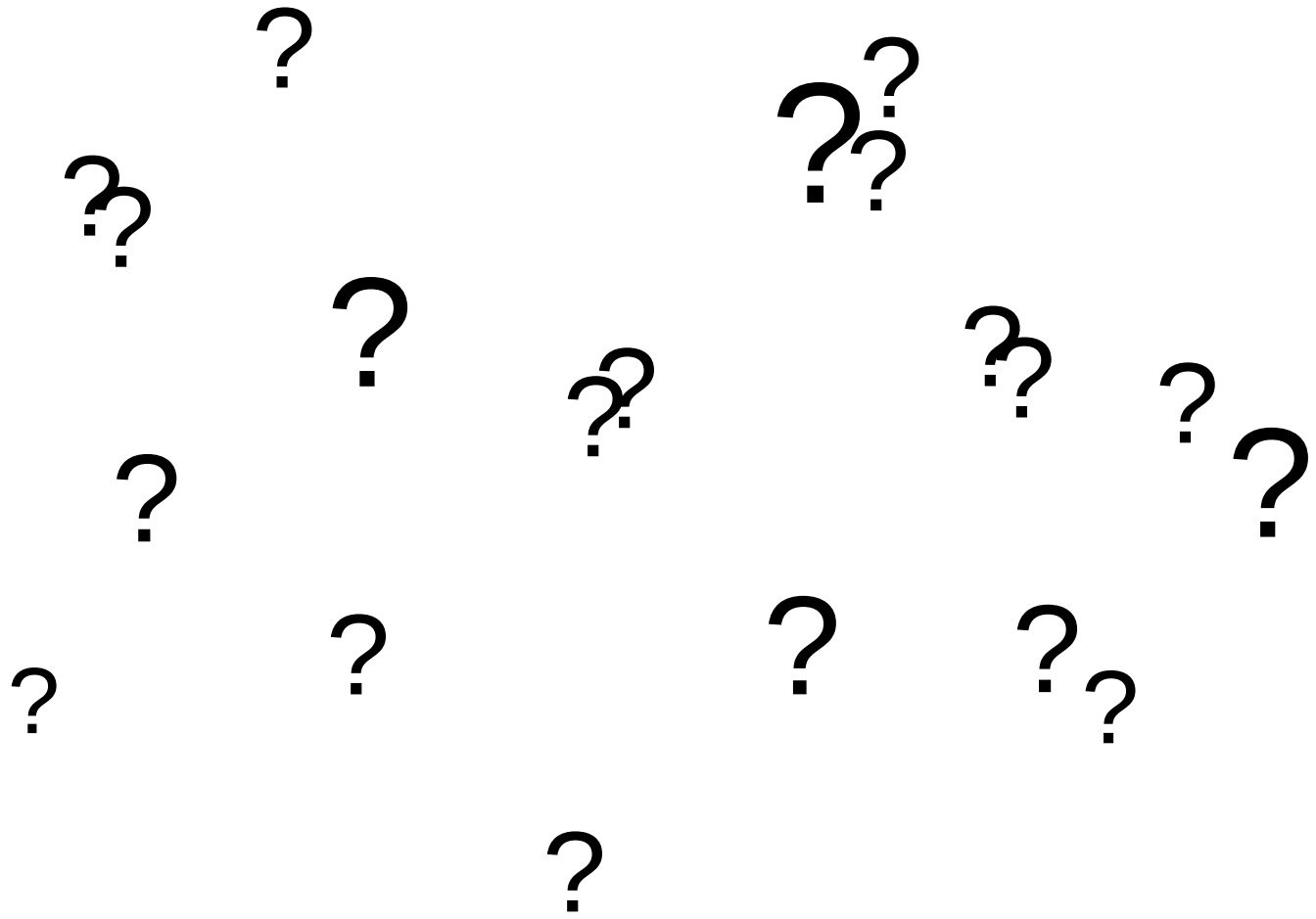
Dr. Benjamin Hofner
Section Biostatistics

Some methods

- Hierarchical testing
  - no α adjustment needed
- α-splitting
  - e.g., Bonferroni (α / # tests)
- Dunnett‘s test
  - multiple comparisons with a single control
- Graphical approaches
  - combination of hierarchical testing with α-splitting and α-recycling methods



Questions? Comments?



Dr. Benjamin Hofner
Section Biostatistics