EU regulatory guidance on multiplicity issues and missing data

Armin Koch
Bundesinstitut für Arzneimittel
und Medizinprodukte
Kurt-Georg-Kiesinger Allee 3
D-53175 Bonn

The views expressed in this paper are those of the author and not necessarily those of the BfArM

History

Following release of ICH-E9 areas of "unmet statistical need" were identified by the Efficacy Working Party (EWP):

- switching between superiority and non-inferiority (2000)
- missing values (2001)
- (1) meta-analysis and (2) one pivotal trial (2001)
- multiplicity issues (2002)
- adjustment for baseline co-variates (2003)
- choice of the non-inferiority margin (2006)

and later:

- "reflection" on adaptive designs in late phases of drug development (2007)

Aim and intended audience

Aim:

Harmonizing the views of the member states in order to avoid discrepant regulatory requirements (and decision making) within Europe

(***but:*** development and discussion of these guidelines also improved dialogue between regulatory and industry statisticians).

Audience:

Medical reviewers in member states (because at that time even less statisticians have been employed by regulatory agencies)

therefore all documents are written in a non-technical style that should be understandable to all parties.

European view:

Two guidelines refer to the topics discussed in this meeting:

Points to consider on multiplicity issues in clinical trials (CPMP/EWP/908/99)

Rapporteur: Prof. Joachim Röhmel, Germany

Points to consider on missing data (CPMP/EWP/1776/99)

(which will be revised in the (near) future)

Rapporteur: Dr. Ferran Torres, Spain

all guidance documents are available from: http://www.emea.europa.eu

PtC on multiplicity issues in clinical trials:
Overview

Objective:

Control of the study-wise rate of false positive conclusions at an acceptable level α is an important principle and is often of great value in the assessment of results of confirmatory clinical trials.

Golden rule:

no choice, no need to adjust the type-1-error.

Obviously,

clinical trials offer many natural opportunities for multiplicity, but beyond this, choice may be introduced in a subtle way…

PtC on multiplicity issues in clinical trials:
Overview

5 key questions:

- Adjustment of multiplicity – when is it necessary and when is it not?
- How to interpret significance with respect to multiple secondary variables and when can a claim be based on one of these?
- When can reliable conclusions be drawn from a subgroup analysis?
- When is it appropriate for CPMP to restrict licence to a subgroup?
- How should one interpret the analysis of “responders” in conjunction with the raw variables?
- How should composite endpoints be handled statistically with respect to regulatory claims?

PtC on multiplicity issues in clinical trials:

Adjustment of multiplicity – when is it necessary and when is it not?

Topics:

- Two or more primary variables are needed to describe clinically relevant treatment benefits, co-primary, or hierarchically ordered;
- Analysis sets (one should be primary, the others are supportive);
- Alternative statistical methods – multiplicity concerns (pre-testing);
- Multiplicity in safety variables (extreme risks and small P-values both can be indicative, adjustment would be counterproductive)
- More than two treatment arms (3-arm GS design, fixed combinations, multiple dose factorial designs, dose-response studies: demonstration of efficacy vs. identification of a dose)

PtC on multiplicity issues in clinical trials:

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

Variables expressing supportive evidence

No claims are intended; confidence intervals and statistical tests are of exploratory nature.

Secondary variables which may become the basis for additional claims

Significant effects in these variables 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.

Variables indicative of clinical benefit

If not defined as primary variables, clinically very important variables (e.g. mortality) need further study when significant benefits are observed, but the primary objective has not been achieved.

PtC on multiplicity issues in clinical trials:

Reliable conclusions from a subgroup analysis and restriction of the license to a subgroup

Summary:

Reliable conclusions from subgroup analyses generally require pre-specification and appropriate statistical analysis strategies. A license 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.

also important:

The evaluation of the uniformity of the treatment effects across subgroups is a general regulatory concern. Some factors are known to cause heterogeneity of treatment effects such as ender, age, …, or differences in absorption or metabolism. Analysis of these important subgroups should be a regular part of the evaluation of a clinical study (when relevant), but should usually be considered exploratory.

PtC on multiplicity issues in clinical trials:

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

Summary:

If the ”responder” analysis is not the primary analysis it may be used after statistical significance has been established on the mean level of the required primary variable(s), to establish the clinical relevance of the observed differences in the proportion of “responders”. When used in this manner, the test of the null hypothesis of no treatment effect is better carried out on the original primary variable than on the proportion of responders.

With some sense of self-criticism:

- very often regulators don't state, what would make a relevant difference in response rates
- it is wise to plan (at least) for a good trend in response rates

PtC on multiplicity issues in clinical trials:
How should composite variables be handled statistically with respect to regulatory claims?

Summary

Usually, the composite variable is primary. All components should be analysed separately. If claims are based on subgroups of components, this needs to be pre-specified and embedded in a valid confirmatory analysis strategy. Treatment should beneficially affect all components, or at least should the clinically more important components not be affected negatively. Any effect of the treatment in one of the components that is to be reflected in the indication should be clearly supported by the data.

Important recommendation:

"Resolution" of the trial (i.e. the ability to exclude differences in e.g. mortality) should be determined upfront.

PtC on multiplicity issues in clinical trials:
Summary (and personal view from today's perspective)

Importance of the topic is well recognized today:

- study planning (in this aspect) is more precise nowadays
- little would need to be changed
- few things would need to be amended
- why Bonferroni Adjustment for primary endpoints was left out (and see: Röhmel et al. Biom. J. (48) 2006, 916 f.)

but:

still it is true that subtle choice is (often un-intentionally) introduced into the design and analysis of clinical trials.

PtC on missing data:
Structure of the document

1. Introduction
2. The effect of missing values on data analysis and interpretation
   - 2.1. Power and variability
   - 2.2. Bias
3. Handling of missing data
   - 3.1. Complete case analysis
   - 3.2. Imputation of missing data (scope, (simple) methods)
4. General recommendations
   - 4.1. Avoidance of missing data
   - 4.2. Relevance of predefinition
   - 4.3. Analysis of missing data
   - 4.4. Sensitivity analysis
   - 4.5. Final report

PtC on missing data:
Practical recommendations

Avoid missing values:
- by design (mortality is a brilliant endpoint, realistic study duration)
- encourage data retrieval after drop-out

Predefine what to do:
- discuss potential reasons for missing values (and what to do)
- discuss why analysis strategy should be conservative
- update during blinded review

Analysis of patterns:
- check for differential (informative) drop-out (frequency, timing)
- investigate baseline of patients with / without missing values

Sensitivity analysis:
- should be used in plural, only

PtC on missing data:
Examples for differential drop-out

Random selection from my reviews:

Comparator	*Treatment 1*	*Treatment 2*	*P-Value*
Active control	308/345	143/172	0,049
Active control	260/326	566/662	0,031
Active control	588/2266	635/2275	0,140
Active control	12/180	5/179	0,084

While it is well possible,

- that drop-out is informative even if there is no indication for differential drop-out,
- once drop-out is differential, this is definitively informative!

PtC on missing data:
Summary (and personal view from today's perspective)

Missing values:

- statistical theory is well developed for MCAR, MAR, and other situations,
- but unfortunately we either don't know or definitely don't believe that these conditions are fulfilled on a regular basis,
- empirical investigations into how much missing values can be tolerated under which conditions are still rare,
- if drop-out is informative, at best, we can try to explore this information,
- PtC: "Missing data violate the strict ITT principle":
  does imputation lead to an ITT-analysis, or do we have to understand "missing" as an outcome?

Finis

In the end missing values are (and will always be) the more challenging topic:

- multiplicity is purely a planning issue,
- but with missing values reality comes into play
  (which loves to ignore all our attempts to plan clinical trials as perfect as possible).

Thanks for your consideration!