EMA Workshop on Multiplicity Issues in Clinical Trials
16 November 2012, EMA, London, UK

(http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/events/2012/06/event_detail_000589.jsp).

Summary
The workshop's topics included usefulness and limitations of newly developed strategies to deal with multiplicity and multiplicity arising from interim decisions. Objectives were to discuss current standards and strategies to address multiplicity in clinical trials and to identify issues where guidance is missing so far. Key opinion leaders from regulatory agencies, academic institutes and industry presented various aspect of multiplicity within a single clinical trial, combination of several trials and the clinical development across different regions.
EFSPI has been invited by EMA to participate.

New sophisticated statistical methods are available to adjust for multiplicity in a correct and efficient statistical way. But that won't solve the whole issue of multiplicity in clinical development. We should realize that analyses are provided to support the complex process of making a decision on the Benefit-Risk in the end. Statistics play a role in making such decisions, but so do biological plausibility and clinical relevancy. Graphical procedures are well appreciated and support transparency and rigor in the planning of clinical trials.

The new guideline could discuss the following topics: (1) role of secondary endpoints and relationship between hypotheses, (2) multi-regional development, (3) reasonable simultaneous confidence intervals corresponding to multiple testing hypotheses, (4) confirmatory conclusions in subgroups (5) multiplicity in parallel studies (in particular in bio-equivalence studies: how may studies were needed)

It became clear to me that a new guidance on subgroup analyses will appear in 2013, but whether there will be a new guidance document on multiplicity in the near future will remain the question.

Egbert Biesheuvel (MSD) on behalf of EFSPI.
(personal view and notes)

<u>Session 1: Experiences with the current guidance document</u>

<u>1. CHMP's view on multiplicity; through assessment, advice and guidelines, Rob Hemmings</u>

- More than one chance to win – to win what? Marketing authorization, success of a trial, or on individual endpoint?
- Important in the multiplicity concept is that the strength of evidence is properly understood (and not only by statisticians)
- Guideline is mainly focused on multiplicity at the level of a trial
- Multiplicity discussion facilitates planning for the sponsor and a more predictable regulatory review, is generally well handled through pre-specs and adjustments, unless …. The trial fails. Rob gave a nice illustration of what he called 'Texas Sharpshooting': shoot at random with a gun on the wall and draw circles around the bullets, with the bullets as bullet eyes… perfect shots!
- New brilliant multiplicity strategies are available on how to do it; but not whether it must (e.g. 2 groups with 1 endpoints: no need) where it is preferable.
- Should considerations of the need to adjust and how to adjust go beyond the data of the trial? It is should it include biological plausibility and the number of pivotal trials needed before getting two significant trials?
- Questions during review: (1) on subgroup populations, (2) different requirements between the regions: write separate SAPs (thus multiple chances), take the most strict one, …(3) "is our complex adjustment appropriate?"- Yes, it is, (4) "is our adjustment necessary?" – very good question!
- Think about could vs should!

<u>2. The FDA perspective, Kathleen Fritsch</u>

- All important "claims" have overall Type I error control. Define claim: 1) indication, 2) any primary /secondary endpoint 3) any statement that appears in the labeling?
-
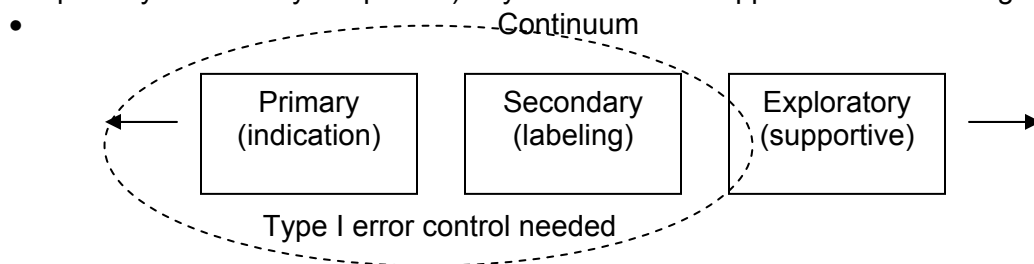
- In general control is needed for primary and secondary endpoints. Secondary could lead to additional labeling claims. Exploratory is anything else: hypotheses generating.
- Permissible for labeling: primary but also secondary endpoints; statistical significant is not sufficient, information on clinically meaningful needed. Can be by description of graphics: time course, subgroups, components of composite endpoints
- Targeted subgroups: want approval for most general population for which the drug is efficacious, this might be different than a statistical significant hypothesis test.
- Additional challenges if the procedures used rely on additional assumptions. For example when is Hochberg's procedure permissible?
- Use of sequential methods: please select only a few secondary endpoints (usually not naturally ordered), the rest is exploratory

- Clear messages:
  - choose "need to have" endpoints and not "nice to have" endpoints. (not only to blame industry, also guilt by regulators)
  - ensure good match between study objectives and multiplicity control and take time to evaluate performance

3. Current experience with multiplicity issues in PMDA, Eisuke Hida

- No guidance document is yet available. Change from avoiding multiplicity issue to consideration to address it in an appropriate way
- Efficiency increasing: multi-regional clinical trials. See Japanese guidance "Basic Principles on Global Clinical Trials" and its "reference cases" (PMDA website ….)
- Topic is complex, therefore important to discuss and share experiences
- Case experiences: (1) disagreement whether first primary endpoint for both high dose and low dose and subsequently the secondary endpoints, or first primary and secondary endpoint for high dose and then low dose, (2) different primary endpoint analysis methods are planned for different regions without multiplicity adjustment. Sometimes inevitable and understandable, but how to handle the overall results in a publication?, (3) please go for prior consultation

4. The update of the multiplicity guideline, Norbert Benda

- Current guidance from 2002 is already comprehensive, so what to enhance?
- New guideline on subgroup analysis is in preparation.
- When do you need confirmatory claims for secondary endpoints, is not addressed. (EU and US have a different environment for "claims"
- Clarification needed for: terminology, role of additional claims, do we need confidence intervals, more complex multiplicity frameworks.
- Clinical assessment often ignores the design " … you clearly see that the high dose is effective…"
- Evaluation of Benefit-Risk profile asks for proper confidence intervals
- Strong FWER or also False Discovery Rate?
- Proposed topics: (1) usefulness and limitations of new strategies, (2) confirmatory conclusions in subgroups, (3) interim decisions, (4) multi-regional developments, (5) simultaneous confidence intervals corresponding to Multiple Testing Procedures
- To resolve: (1) role of secondary endpoints and relationship between hypotheses, (2) transparency (3) reasonable CIs and B-R, (4) multiplicity in parallel studies (in particular in bio-equivalence studies: how may studies were needed)

Discussion

- Rob: most common question/issue is on more clarity on claims. Differences in secondary endpoints between EU & US (being more structural in the US). It is clear if the company is in discussion with FDA. How sensible is the hierarchy from a pharmacological point of view should play a role.
- Kathleen: clinicians want to see more and more endpoints, but please focus on what you need to know.

- Norbert: often questions like "would you accept the following procedure ….". In particular for secondary endpoints it remains problematic to assess.
- Industry: is statistics not going too far? The clinicians don't understand it anymore. Reaction Peter Bauer: every statistical detail that is not well understood within a minute by others is rubbish is unacceptable. This also doesn't hold for other disciplines.

Session 2: Usefulness and limitations of newly developed strategies to deal with multiplicity Part 1

1. Multivariate Analysis of treatment in Multiple Sclerosis using the Wei-Lachin procedure, Thomas Zwingers

- Combines equally important variables of different nature in one test with directional alternatives. Is the non-parametrical equivalent of Hotelling's T test.

Discussant: why not seen earlier and more often in submissions? It might there are not so many situations with equally important endpoints.

2. Dunnett and Bonferroni Corrections in Bioequivalent Testing, Jiri Hofman.

- Example of a 5-way cross-over design with more than two formulations.
- Dunnett: the 'Guideline on the investigation of bioequivalence' CPMP/EWP/QWP/1401/98 Rev 1/corr, seems to mention that each formulation should be compared independent from each other.
- Two one-sided t-tests translate into confidence intervals, which are not symmetrical.

Discussant: (Remark Martin Posch) in fact these are not simultaneous confidence intervals, but could only help you to reject the correct hypotheses.

3. Optimal multiplicity adjustment and the necessity to use separable multiple test procedures as gate keeper for secondary endpoint testing: case study, Vincent Haddad.

- Three arm study with Overall survival as primary endpoint and Progression Free Survival as secondary endpoint.
- Only "separable" Multiple Testing Procedures could open the gate to secondary endpoints.
- Different strategies for phase III have been considered (Bonferroni: low power; Hochberg only: no secondary endpoints; sepr. mixture (gatekeeping): more complex).
- Original plan to use Hochberg only was acceptable at scientific advice.
- What would EMA recommend?

Discussant (Franz Koenig): showed that the proposed method did not control the FWER is the strong sense. A guideline will never be able to include all up-to-date methods. (Remark Rob Hemmings): Why interest to control the error rate for PFS (secondary) if OS is the primary endpoint?

4. Multiplicity issues in defining the testing strategy for two large outcome studies, Jennifer Shannon

- Discusses two major cardiovascular trials with composite primary endpoints and its pre-planned integrated analyses.
- To address the multiplicity issues within each study and the integrated analysis, a hierarchical testing with a gate keeping strategy was proposed.

Discussant: (Remark Armin Koch) if only one study has primary endpoint significant and the other not, then the regulatory requirements are not met. You need two self-standing trials to see consistency, not one stand alone trial and a meta-analysis.

Session 3: Implications of multiplicity for estimation

1. Multiplicity and Estimation, Peter Bauer

- Selection bias: by selection of best treatment (Putter and Rubinstein 1968)
  - Let k = number of treatments, r = information time.
  - No bias by random selection of treatment ($r=0$); in sharply increases with r and is largest with $r=1$ (i.e. post trial selection): So Adaptive Design is less dangerous! (but earlier selection increases the wrong selection)
  - $\sqrt{MSE}$ does not increase with k to the same extent as the bias, $\sqrt{MSE}$ increases close to linear for r.
- Reporting bias: is in general negative
  - If effect at Interim Analysis is large, then selected: regression to the mean
  - If effect at Interim Analysis is small, not selected: stay with this 1st stage effect
- Admission Bias: with two trials; only estimates are reported if both are significant.
- Methods to reduce bias (from sequential trials), shrinkage estimators
- Questions with bias correction: (1) bias or mean square error, (2) what is a suitable criterion for a "good" estimate?, (3) should we report conventional estimators and would regulators agree?, (4) should bias-adjusted estimator be given in the spirit for sensitivity analyses

Discussion:
- Simultaneous confidence intervals exist for certain situations (Hsu 1994) and depend on the data
- More informative confidence sets reduce power: is not acceptable for patients to reduce benefit for this reason
- Do unadjusted confidence intervals have a role to play? (Like unadjusted p-values together with adjusted p-values)
- Remark Peter: in scientific papers a move from p-values to CIs, in regulatory setting still focus on significant p-value
- Decision making is a complex process and not too transparent. Statistical significance will not replace but facilitate.

Session 4: Usefulness and limitations of newly developed strategies to deal with multiplicity Part 2

1. Gatekeeping strategies in Phase III clinical trials with multiple endpoints and doses, Alex Dmitrienko

- Case study based on the Lurasidone development program in schizophrenia
- Principles to follow:
    - Incorporate logical relationships (A) [clinical]
    - Utilize available distributional information (B) [statistical]
    - Select optimal procedure (C) [simulation, power, consistency]

2. Novel multiple testing procedures for structured study objectives and families of hypotheses, Guenther Mueller-Velten

- High, medium and low dose, primary endpoint is CV death, MI and stroke, secondary endpoints are extended composite endpoint and new onset of Type II diabetes.
- "alpha propagation" through weights and "successiveness": only 2nd endpoint if associated 1st endpoint is rejected
- Importance of graphical techniques
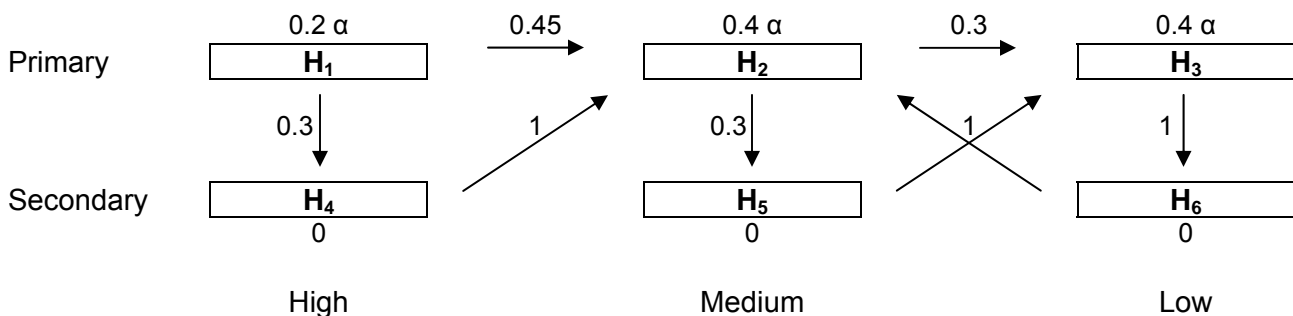
- Secondary: two hypotheses for both endpoints using Bonf-Holm
- EMA could provide in its new guidance a harmonized terminology and framework categorizing study objectives and endpoints for their impact on approval and labeling respective need for Type I error control.

3. Multiplicity: Is it a value to make it so complicated? Andy Stone

- Should we do it, rather than how to do it. And if we have to, we will even if complicated
- Strong control is in danger to become self-defeating
- Given the complexity, does it add to our assessment of medicines? Analyses are provided to inform prescribers the nature of Benefit-Risk.
- How robust are our complicated methods? Multiplicity is closely related to reproducibility. Two pivotal trials. We should not forget the scope of the development program.

<u>Discussion</u>

- Rob Hemmings: Assume multiple Phase III trials (say 3) with a complex process to control for multiplicity. What if each trial stops at different endpoints? On the other hand, you have 3 independent large studies. Look at the whole richness of all data.
- Armin Koch: should there be more incentive to go into phase III with more dosages just because new complex methods are available? Frank: no but in case you are in that situation, new methods do exist.
- Willy Maurer: Cox regression is also complex, but can be used and is accepted without explaining all details.
- Peter Bauer: Why can't the statistician use complex methods to do his job and to support the complex process of making a decision? (Statisticians don't play an important role in making such decisions)
- The graphical procedures are well appreciated and help the clinician to explain their reasoning in the development process. It supports rigor in the planning.
- Strong FWER control over the primary and secondary endpoints will at least help the discussion to have to many secondary endpoints. However, there could be other ways to get there: just mention a maximum of 3 secondary endpoints in the guideline.