Multiplicity in Drug Development

Søren Lophaven

Types of multiplicity

- Multiple endpoints
- Multiple treatments or doses
- Multiple looks (interim analyses, adaptive design, sequential design)
- Multiple trials

Types of multiplicity (ICH-E9)

". ... Multiplicity may arise, for example, from multiple primary variables (see Section 2.2.2), multiple comparisons of treatments, repeated valuation over time and/or interim analyses (see Section 4.5) ...."

"When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data may necessitate an adjustment tot he type I error. ..."

What's the issue with multiplicity?

- For each test the cumulated risk of type I error increases
- Something must be done:
  - Split alpha (Bonferroni-style)
  - Hierarchy of tests

Multiplicity

- If only one null hypothesis is tested then the probability of a false positive is 5%
- If two independent null hypotheses are tested then the probability of at least one false positive is $\sim$ 10% $(1 - 0.95^2)$
- If five independent null hypotheses are tested then the probability of at least one false positive is $\sim$ 23% $(1 - 0.95^5)$

Multiple primary endpoints

- If two or more primary endpoints are needed to claim clinical relevant benefits:
  - No formal adjustment is necessary
  - All H_0 need to be rejected
- If two or more primary endpoints are ranked according to clinical relevance:
  - No formal adjustment is necessary
  - No confirmatory claims can be based on endpoints having a lower rank than hypothesis first not be rejected
  - Hierarchy of tests

Multiple secondary endpoints

- No claims intended:
  - No adjustment necessary
- Additional claims intended:
  - Primary objective need to be achieved
  - Hierarchically ordering of test is a valid procedure
  - Primary and secondary endpoints part of same overall hierarchical testing strategy

More than two treatment arms

- Control of type I error necessary
- Many types of designs
- Split of alpha to adjustment for multiplicity could be necessary depending on design:
  - Bonferroni: (alpha level) = (5%)/(#tests)
  - Dunnet's test: Different treatments all compared to the same comparator or placebo. Hence a slightly higher level than Bonferroni

Testing strategy in a depression trial

- Bonferroni: p = 0.05/2 = 0.025
- Testing hierarchic

	Low dose vs pbo p=0.025		High dose vs pbo p=0.025	
Endpoint	P-value	Succes	P-value	Succes
MADRS	0.015	Yes	0.010	Yes
CGI-S	0.029	No	0.013	Yes
PSQI	0.021	No	0.021	Yes
MEI	0.06	No	0.041	No

Can we avoid dealing with it?

EMA - Points to consider on multiplicity issues in clinical trials:

"A clinical study that requires no adjustment of the type I error is one that consists of two treatment groups, that uses a single primary endpoint, and has a confirmatory statistical strategy that pre-specifies just one single null hypothesis relating to the primary endpoint and no interim analysis."

Are there ways to reduce the issue?

ICH-E9:

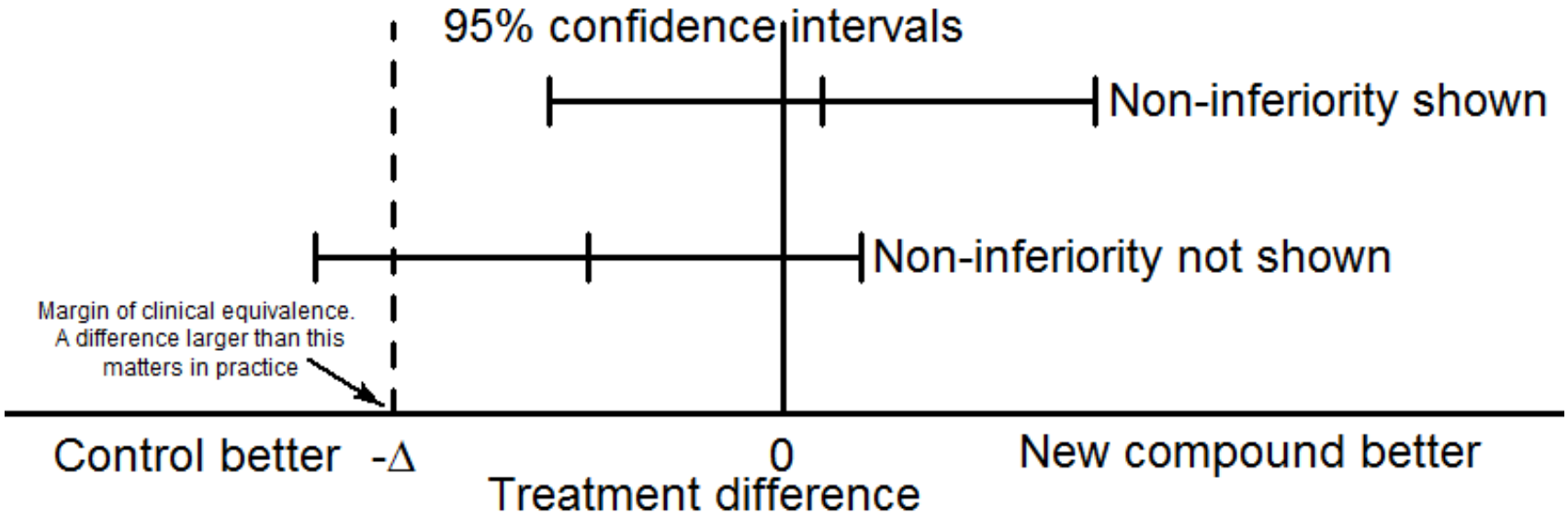
"... Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), the use of a summary measure such as 'area under the curve' (repeated measures). ..."

Multiple trials

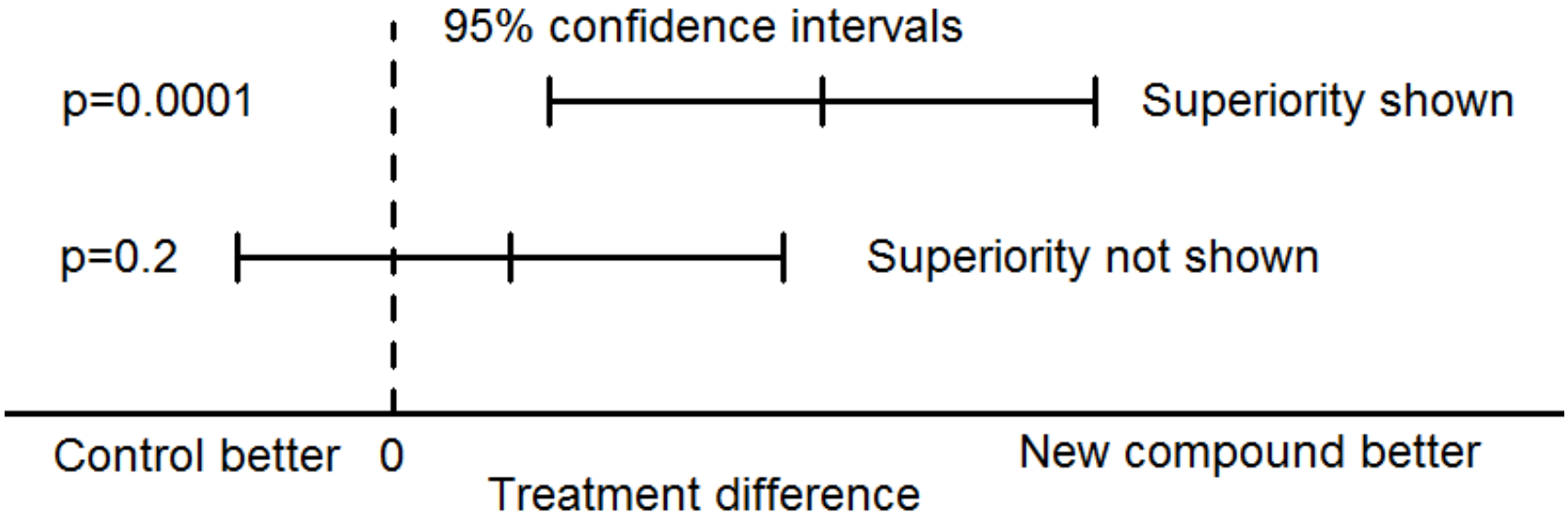
- Asking for two significant (α=0.05, two-sided) out of two pivotal trials (FDA) is a multiple test. If there is no treatment effect at all the chance of a false approval is $0.025 \times 0.025 = 0.000625$
- If more than two "pivotal" trials are run with ineffective treatments the probability of getting at least two significant trials increases, e.g. for 4 trials to $6 \cdot 0.025 \times 0.025 \times 0.975 \times 0.975 = 0.00363$
- If approval is based on a single trial (one pivotal trial) a very low significance level has to be applied to get the same overall false approval rate

Switching between superiority and non-inferiority

- Design trial to show non-inferiority
- Hierarchical testing procedure $\Longrightarrow$
- If non-inferiority has been shown then superiority can be tested with adjustment of type I error



Switching between superiority and non-inferiority



Other methods for dealing with multiplicity

- Start with the smallest p-value of the k tests. If the null hypothesis is rejected at the level α/k compare the second smallest p-value with $\alpha/(k-1)$; after rejection, increase the level to $\alpha/(k-2)$, ... (Holm, 1979). STEP-DOWN

- Start with the largest p-value. If it is smaller than α reject all null hypothesis. If not, look if the second largest p-value is below $\alpha/2$ then reject the remaining k-1 null hypothesis. If not, compare the third largest pvalue with $\alpha/3$, ... (Hochberg, 1988; Hommel, 1988). STEP-UP