Key comments from the public consultation and the regulatory challenge

Hannover Medical School

Armin Koch, DE SAWP and BSWP member
Institut für Biometrie
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625 Hannover

A Guideline on Subgroups in Phase III Clinical Trials

Paradigm of phase III clinical research:

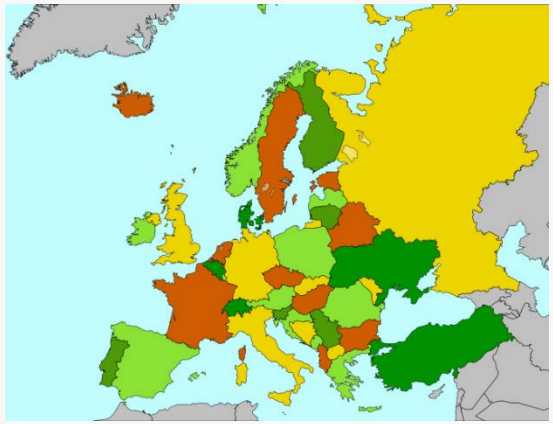
Trials should not fine-tune the patient population unless there is a clear rational.

Prize to be paid:

Homogeneity of the treatment-effect in relevant subgroups of the patient population is non-trivial and needs to be verified.

Relevant subgroups:

Something that needs to be defined, but demography, gender, disease characteristics, co-medication, center, region and country are plausible candidates



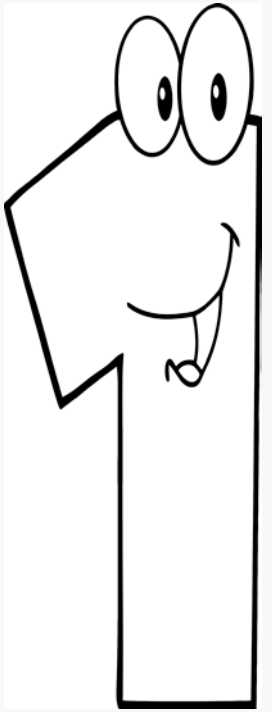
Did the importance of subgroups change?

Medics will say no, because they were always interested in subgroups.

However, standards of evidence have changed:

- in former times (two-trials rule of the FDA) we had 2 (usually PBO controlled) studies in the US ***and*** 2 (usually active controlled) studies in the EU.
- nowadays assessment of efficacy and benefit/risk is based on one world-wide pivotal study planned with an adaptive design intended to justify licensing in all the ICH-regions.

☞ If consistency / replication is considered important, nowadays assessment needs to be done *within* instead of *between* studies.



Comments to the guideline

Guideline has been under consultation from the Feb to July 2014.

19 persons / parties provided comments summing up to 150 pages.

Full representation of:

- Industry and interest groups,
- Academia and learned societies
- HTA bodies (IQWIG, NICE)
- EORTC
- Individual parties with interest in the topic

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

7 August 2014
EMA/477011/2014
Committee for Medicinal Products for Human Use (CHMP)

Overview of comments received on 'Guideline on the investigation of subgroups in confirmatory clinical trials' (EMA/CHMP/539146/2013)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	Teva Pharmaceutical Ltd
2	Stallergenes, Policy and Regulatory Intelligence department
3	PSI/EFSPI
4	International Society of Clinical Biostatistics (ISCB)
5	IFAPP (International Federation of Associations of Pharmaceutical Physicians)
6	Stephen Senn, Franz König, Ralf-Dieter Hilgers and Geert Molenberghs on behalf of the IDEAL team
7	Ian White (Medical Research Council Biostatistics Unit, Cambridge, UK)
8	German Region of the International Biometric Society (IBS-DR) and German Society for Medical Informatics, Biometry and Epidemiology (GMDS)
9	Erik Cobo, Barcelona Tech, Statistics and Operational Research department
10	EFPIA – Pär Tellner
11	ISCB Member – Nazneen Shariff
12	Technology Appraisals Programme – National Institute for Health and Care Excellence (NICE)
13	F. Hoffmann-La Roche Ltd.
14	European Organisation for Research and Treatment of Cancer (EORTC)
15	Willi Sauerbrei (Center for Medical Biometry and Medical Informatics, Medical Center – University of Freiburg, Germany) and Patrick Royston (MRC Clinical Trials Unit at UCL, London, UK)
16	Pfizer Inc
17	James P. Scanlan, Attorney at Law, Washington, DC, USA
18	CBG-MEB
19	Ralf Bender (IQWiG, Germany)

30 Churchill Place • Canary Wharf • London E14 5EU • United Kingdom
Telephone +44 (0)20 3660 6000 Facsimile +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

An agency of the European Union

Main topics of the discussion

1. Clearly state the objective / goal:
   to pay the prize: assessment of the trial after confirmatory testing has ended; critically challenging the primary outcome and conclusion of the trial.
2. The concept of 3 levels of subgroup analyses and the wish to limit the number of subgroup-analyses.
3. Be [more] specific about [particularly Bayesian] methodology.
4. Need to distinguish qualitative / quantitative heterogeneity?
5. The role of heterogeneity testing, definition of consistency.
6. Terminology (heterogeneous population, inconsistent estimates)
7. Need to distinguish discussion for superiority and non-inferiority trials?
8. Powering for subgroup effects?

Main topics of the discussion

9. Modelling vs. subgroup analyses, continuous vs dichotomized variables, risk of misclassification;
10. Be more specific regarding implications for the label and general implications of differential benefit/risk in subgroups.
11. The “You may be mislead!” - discussion

12. Use information from subgroups as a rich source of information

Regulatory mandate: the precautionary principle

An overall positive treatment effect may be put into perspective in subgroups by:

- no effect in relevant subgroups of the patient population
- indication of harm
- negative benefit/risk in subgroups
- substantial heterogeneity

Please note:

Sometimes it is a good thing, if a subgroup can be identified that reliably bears a risk and can thus be excluded (cf. Rofefcoxib (VIOXX) as an example, where this hasn’t been possible).

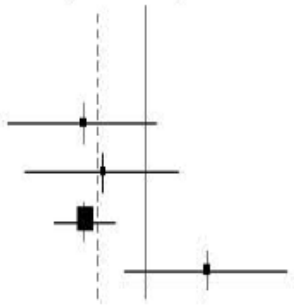
It is difficult to ignore evidence of negative benefit/risk.

“Heterogeneity” may be informative

The Plato trial, comparing Ticagrelor to Clopidogrel in 18,000 patients with ACS demonstrated superiority, but regional differences became obvious from the results (p_{Het}~0,05).

Is it wise to pretend that this is an American problem?

Figure 20 Forest Plot: Results by Region (K-M)



Characteristic	Total Patients	KM % at Month 12 Ti.	KM % at Month 12 Cl.	HR (95% CI)
Region				
Asia/Australia	1714	11.4	14.8	0.80 (0.61, 1.04)
Cent/Sth America	1237	15.2	17.9	0.86 (0.65, 1.13)
Euro./Mid E./Afr.	13859	8.8	11.0	0.80 (0.72, 0.90)
North America	1814	11.9	9.6	1.25 (0.93, 1.67)

The issue of heterogeneity testing

Limitations of the “classical” heterogeneity test are well known:

Table 1: empirical type 1 error / power in a model with two FEMs in two strata

Situation	ratio	I^{2*}	Cochran’s Q	Q-Rule	G-Rule	E-Rule	KI-Rule
H_0	50:50	0.1489	0.1449	0.1427	0.0723	0.0179	0.0446
H_{1G}	50:50	0.4192	0.4804	0.4798	0.4580	0.2232	0.2998
H_{1E}	50:50	0.6136	0.7192	0.7172	0.7006	0.4395	0.5656
H_0	70:30	0.1495	0.1494	0.1541	0.2014	0.0725	0.2071
H_{1G}	70:30	0.3093	0.3472	0.3480	0.5110	0.3321	0.4299
H_{1E}	70:30	0.4374	0.5018	0.5027	0.6753	0.4927	0.6003
H_0	90:10	0.1486	0.1474	0.1459	0.4580	0.2237	0.5127
H_{1G}	90:10	0.2183	0.2356	0.2370	0.6176	0.4545	0.6100
H_{1E}	90:10	0.2668	0.2909	0.2943	0.6831	0.5496	0.6742

*

Why not use it in a way medics would use a blood-count?

False positives should be no problem, if there is an agreement that this is signal detection.

Torn between two extremes

The subtle balance between:

increasing the type-1-error by means of multiple testing in subgroups

and

overlooking important untoward effects in subgroups

can only be ameliorated by means of pre-planning and specification of what is a relevant subgroup at the planning stage.

In this, statisticians fear eventually too much to be mislead by (good quality data).

Under which conditions could a subgroup finding be convincing?

In case the overall trial is not significant usually from statistical grounds no further confirmatory testing is possible (type I error is exhausted).

Any step further can only be based on a case by case decision.

Most important point:

- a generally acceptable argument should exist, why straightforward replication is not possible,

because,

- replication of promising subgroup-findings in an independent trial

is standard if trial is overall not significant, but subgroup findings suggest efficacy at least in parts of the patient population.

Under which conditions could a subgroup finding be convincing?

Criteria have been (repeatedly) presented in the literature:

Ten criteria used to assess credibility of subgroup effect

Design

- Was the subgroup variable a baseline characteristic?
- Was the subgroup variable a stratification factor at randomisation?*
- Was the subgroup hypothesis specified a priori?
- Was the subgroup analysis one of a small number of subgroup hypotheses tested (≤5)?

Analysis

- Was the test of interaction significant (interaction P<0.05)?
- Was the significant interaction effect independent, if there were multiple significant interactions?

Context

- Was the direction of subgroup effect correctly prespecified?
- Was the subgroup effect consistent with evidence from previous related studies?
- Was the subgroup effect consistent across related outcomes?
- Was there any indirect evidence to support the apparent subgroup effect—for example, biological rationale, laboratory tests, animal studies?

*Item was not included in our previously published list of criteria for subgroup credibility

Sun, X. et al: Credibility of claims on subgroup effects in randomized controlled trials (BMJ 2012 (344)).

Under which conditions could a subgroup finding be convincing?

Guiding principles for this case-by-case decision include:

- a pharmacological rational, or a mechanistically plausible explanation, should at best exist for differential treatment effects in subgroups,
- a priori, or external evidence should exist that subgroup is well known,
- stratification of the randomisation as an indicator,
- convincing P-value (not borderline in a borderline trial)
- the overall outcome of the trial should at a minimum substantiate the claim that no harm is introduced by the experimental treatment. It is not possible to claim treatment benefit in one subgroup, if another subgroup suggests that also harm may be introduced w/o replication,
- good overall safety and subgroup safety, or convincing benefit/risk assessment from subgroup is possible
- substantial heterogeneity?
- replication?

Finis:

- one needs to be very brave if one wishes to leave the arena of pre-planned decision making with full control of the type-1-error
- a bridge has been built
- P<0.05 is no longer the criterion, epidemiology-style decision making needed
- better methodology needed, as well. Guideline will not mandate specific methodology, but mention the required information for decision making.